

6th annual

TMLS 2022

The Carlu, Downtown Toronto

TORONTO MACHINE LEARNING SUMMIT

*Passes give access to any session. Build your own program,
your way.*



All-New Workshops, Case-
Studies, Best Practices &
Lessons Learned

Nov. 28th – 30th

www.torontomachinelearning.com



6th annual

TORONTO MACHINE LEARNING SUMMIT 2022

This November, Toronto will be home
to 80+ Learn Technical and Strategic Leaders
from



Nov. 28th–30th

www.torontomachinelearning.com

Longer form, many hands-on, with instructions upon sign-up



9: 00 AM	Hands-on Workshop	Introduction to Kubernetes for MLOps <i>Mohammed Sabri, Chief MLOps Officer Asim Sultan, MLOps Engineer, Rocket Science Development</i>
9: 30 AM	Hands-on Workshop	Open Source Intelligent Application Delivery on Kubernetes <i>Arthur Vitui, Senior Data Scientist & Solution Architect Specialist, RedHat Canada</i>
9: 30 AM	Hands-on Workshop	Graph Neural Network Modeling in Drug Discovery using PyTorch <i>Dr. Nasim Abdollahi , Postdoctoral Fellow, ML Researcher, University of Toronto, Dr. Farnoosh Khodakarami, Computer Scientist and ML Researcher, Cyclica</i>
10 : 00 AM	Workshop	Demystifying De-Identification <i>Patricia Thaine, Co-Founder & CEO, Private AI</i>
12: 00 PM	Hands-on Workshop	Graph ML – The Next Level of Machine Learning <i>Jörg Schad, CTO, ArangoDB</i>
12: 00 PM	Deep-dive	Introduction to NLP & a Step by Step Implementation of a Real world Use Case from TELUS <i>Mahmudul Hasan, Lead Data Scientist, TELUS Business Marketing</i>
12: 00 PM	Hands-on Workshop	Beyond the Basics: Data Visualization in Python <i>Stefanie Molin, Software Engineer / Data Scientist/ Author, Bloomberg</i>
2: 30 PM	Hands-on Workshop	Time Series Anomaly Detection with Machine Learning <i>Angeline Yasodhara, Benjamin Ye, Research Scientists, Georgian</i>
3: 00 PM	Workshop	NLP for Healthcare: Challenges with Processing and De-identifying Clinical Notes <i>Chloé Pou-Prom, Data Scientist, Vaakesan Sundrelingam , Data Scientist, Unity Health Toronto</i>
3: 30 PM	Hands-on Workshop	Distributed Training with PyTorch <i>Shagun Sodhani, Research Engineer, Meta AI</i>
5: 00 PM	Workshop	Train Your Models Faster by Learning How to Profile and Apply System-Level Optimizations <i>Akbar Nurlybayev, VP of Engineering Xin Li and Yubo Gao, Research Engineers, CentML</i>
5: 30 PM	Hands-on Workshop	Iterating on NLP models from R&D to Production <i>Denys Linkov, ML Lead, Voiceflow</i>




[Register here](#)

Tickets give access to all sessions. Build your own program, your way.

November 29th: In Person, Toronto Canada

8:45 am	Registration and Exhibits Open		
9:35 am	Opening Welcome: David Scharbach, Executive Director, <i>TMLS</i>		
9:55 am	<div>Ignite Lightning Keynotes: Wealthsimple, Pachyderm Superwise, SpotHero;</div> <div>What To Look For In Your Next ML Pipeline Deploying a Machine Learning Model in under 15 Minutes at Wealthsimple 7 Questions for Data Scientists Retraining Won't Fix Your Model (Always)</div>		
	10:25 am Break		
10:55 am	<div>Scaling Advanced Analytics in the Worst Crisis in the Industry</div> <div>Nicolás Venegas Oliva, Technical Lead of Advanced Analytics Cristobal Guzman, Staff Data Scientist LATAM Airlines Group</div>	<div>Saving Lives with ML: Applications and Learnings</div> <div>Muhammad Mamdani, VP, Data Science and Advanced Analytics; Director: Temerty Centre for AI Research and Education in Medicine; Professor Unity Health Toronto, University of Toronto</div>	<div>Where's the Road? The Challenge of Autonomous Driving Perception in Winter</div> <div>Steven Waslander, Professor, Director Toronto Robotics and Artificial Intelligence Laboratory (TRAILab), University of Toronto</div>
11:35 am	<div>Text Generation Using Large Language Models - Tips and Tricks</div> <div>Lightning Talks with Open Group Discussion</div> <div>Nikhil Verma, Lg Toronto AI Research Lab, LG Electronics AI Lab He (Richard) Bai, University of Waterloo Ali Sabet, Cohere AI</div>	<div>From Silo to Collaboration - Building Tooling to Support Distributed ML Teams at Twitch</div> <div>Shiming Ren, Sr. Engineering Manager; Safety, MLOps & Chen Liu, Sr. Engineering Manager, Infrastructure & Personalization and ML Infra Twitch</div>	<div>Saga: Continuous Construction and Serving of Large Scale Knowledge Graphs</div> <div>Ihab Ilyas,Professor, Head of Apple Knowledge Platform, ACM Fellow, IEEE Fellow Apple, University of Waterloo</div>
	12:05 pm Lunch Break Exhibition Brain-Dates via Whova App		
1:15 pm	<div>The Application of Mobile Location Data for Vending Machine Site Selection and Revenue Optimization</div> <div>Nikita Medvedev, Winston Li Director of Advanced Analytics, Founder, Coca-Cola, Arima</div>	<div>Navigating the Tradeoff Between Privacy and Fairness in ML</div> <div>Jesse Cresswell, Senior Machine Learning Scientist Layer 6, TD Bank</div>	<div>Declarative Machine Learning Systems: Ludwig & Predibase</div> <div>Piero Molino, CEO Predibase, Inc.</div>
1:50 pm	<div>The Risks of Excluding the Disengaged from your Dataset</div> <div>Danielle Goldfarb, Global Affairs / Economic Policy Researcher RIWI (Real-Time Interactive Worldwide Intelligence)</div>	<div>Real-time Machine Learning: Architecture and Challenges</div> <div>Chip Huyen, CEO, Author Claypot AI</div>	
	2:20 pm Break Exhibition Brain-Dates via Whova App		
2:55 pm	<div>Bias and Relevance in Application of ML in Healthcare</div> <div>Shazia Akbar, Lead Machine Learning Engineer Ali Madani, Director of Machine Learning Shiva, Santosh and Javier Altis Labs, Cyclica, Pfizer</div>	<div>Optimal Beer Pricing: An Optimization Layer for Price Elasticities</div> <div>Eric Hart , Staff Data Scientist Anheuser-Busch</div>	<div>Outracing Champion Gran Turismo Drivers with Deep Reinforcement Learning</div> <div>Varun Raj Kompella Senior Research Scientist Sony AI</div>
3:30 pm	<div>Sensitivity & Interpretability of AI-Models</div> <div>Hanieh Arjmand, ML Researcher Spark Tseung, Applied Data Scientist Lydia.ai</div>	<div>How Freshbooks Built a Fully Automated ML Platform Using Kubeflow; A Declarative Approach to Development of End-to-end ML Pipelines.</div> <div>Valerii Podymov et al., Lead Data Scientist Roshan Issac, Machine Learning Engineer Vlad Ryzhkov, Senior Data Engineer Joey Zhou, Senior Data Engineer FreshBooks</div>	<div>PyScript for Data Science</div> <div>Sophia Yang Senior Data Scientist Anaconda</div>
	Social Events		
	<div>Day 1 Program Ends Networking</div> <div></div>		
4:15 - 5:30 pm	<div>Day 1 Wrap- up & Happy Hour Social</div> <div>* ML Trivia Nintendo Gaming Station * Drinks & Food Prizes & more!</div> <div></div>		

November 30th: In Person, Toronto Canada

	Registration and Exhibits Open		
9:00 am			
	Opening Welcome: Suhas Pai, 2022 Chair and CTO <i>Bedrock AI</i>		
9:35 am	<i>Is AI Being Sufficiently Democratized?</i>		
	Ignite Lightning Keynotes:		
9:45 am	Loblaw, Autodesk, Georgian, Vian AI, Arize AI, Rockwell Automation <ul style="list-style-type: none">Recommender Systems at Loblaw DigitalThe Transformative Role of AI/ML in Heavy IndustriesAre AI Scientists the key to Organization's AI Success?<ul style="list-style-type: none">Monitoring Unstructured Models in ProductionAI & Sustainability: A \$50 trillion Opportunity		
	10:25 am Break		
10:55 am	<i>The Framework for Great ML Products</i> Liran Hason, CEO Aporia	<i>Transforming The Retail Industry with Transformers</i> Rohit Saha, Applied Research Scientist, Kyryl Truskovskyi ML Engineer Georgian	<i>Artificial Intelligence & Digital Pathology; Making the Most of Limited Annotated Data</i> Anne Martel, Professor, Tory Family Chair in Oncology, Faculty Affiliate, University of Toronto, Sunnybrook Research Institute, Vector Institute
11:35 am	<i>Image Generation Using Stable Diffusion - Tips and Tricks</i> Jordan Shaw Half-Helix	<i>Scaling & Evolving the Machine Learning Platform at DoorDash</i> Hien Luu, Head of Machine Learning Platform DoorDash	<i>Interpretability Tools are Feedback Loops</i> Jinen Setpal, ML Engineer DagsHub
	12:05 pm Lunch Break Exhibition Brain-Dates via Whova App		
1:15 pm	<i>Stanford AI Index: 2023 Economic Trends in AI</i> Nestor Maslej Research Manager, Stanford Institute for Human-Centered Artificial Intelligence (HAI)	<i>Marketplace Experimentation in Live Environments at SpotHero: Design, Methods, Lessons Learned</i> Amish Popli Data Scientist, SpotHero	<i>The Role of Alternative Data in Investing</i> Serena McDonnell, Lead Data Scientist Delphia
1:50 pm	<i>Panel: Managing AI and the Associated Risks in a Regulatory Environment</i> Susie Lindsay, Counsel, Law Commission of Ontario Fion Lee-Madan, Co-Founder, COO Sgt. David Cobey, RCMP	<i>Assessing Alignment of Climate Disclosures Using NLP for the Financial Markets</i> Quoc Tien Au & Aysha Cotterill, Data Scientist, Data Analyst Manifest Climate	<i>Latent User Intent Modeling in Recommender Systems</i> Bo Chang, Software Engineer Google Brain
2:25 pm	<i>Panel: Data Fails!</i> Sarah Sun, Director Data Science , Scotiabank Nicolás Venegas Oliva, Technical Lead of Advanced Analytics LATAM Airlines David Scharbach, Executive Director, TMLS	<i>Solving Product Substitutions, the #1 problem in Grocery e-Commerce, through Self-Supervised ML</i> Jawad Ahmed Staff Applied Scientist, Loblaw Digital	
3:00 pm	Closing Remarks		
5:15 - 7:30 pm	Career Fair & Expo 		

[Register here](#)

NOVEMBER 28th WORKSHOPS

9:00 AM
-
1:00 PM EST

Introduction to Kubernetes for MLOps
*Eric Hammel, MLOps Engineer I Mohamed Sabri, Chief MLOps Officer, **Rocket Science Development***

About the Speaker:
A resourceful professional able to bridge skills between Data Science and Infrastructure (Cloud and HPC) to deliver valuable solutions. With experience in prototyping, deploying, and monitoring distributed workloads to drive an organization in translating real-life business problems into scalable data science solutions to generate value.

Mohamed Sabri is a results-driven Machine learning Engineer and MLOps specialist with 8 years of experience in Machine learning and data field. This Trilingual consultant has a very wide range of project experience for various industries. Mohamed Sabri has been as well a director of data science and AI in a startup in Montréal, with his experience on the field and as a manager Mohamed is capable of supporting all organization in their AI project implementation. Mohamed is also the author of the book “Data Scientist Pocket Guide” and instructor and mentor at the University of Austin Texas.

Which talk track does this best fit into?
Workshop

Technical level of your talk?
(Technical level: 5/7)

What you’ll learn:
The participants will get a crash course about Kubernetes and Cloud Native concepts. They will learn how to deploy an application on a managed kubernetes cluster.

Abstract of Talk:
Have you ever wondered what kubernetes and Cloud Native applications are?
Here is the perfect opportunity to get exposed to these complex yet powerful tools & conecepts.
You will discover Container Orchestration, Cloud Native applications, Kubernetes, and application deployment.

9:30 AM
-
11:30 PM EST

Open Source Intelligent Application Delivery on Kubernetes
*Arthur Vitui, Senior Data, Scientist Specialist Solution Architect, **RedHat Canada***

About the Speaker
Arthur is a senior data scientist specialist solution architect at RedHat Canada where with the help of open source software is helping organizations develop intelligent application ecosystems and bring them into production using MLOps best practices.
He is also pursuing his Ph.D. degree in Computer Science at Concordia University, Montreal, Canada, and he is a research assistant in the Software Perfomance Analysis and Reliability (SPEAR) Lab.
His research interests are related to AIOps with focus on performance and scalability optimization.

Which talk track does this best fit into?
Workshop

Technical level of your talk?
(Technical level: 4/7)

Are there any industries (in particular) that are relevant for this talk?
Computer Software, Hospital & Health Care, Information Technology & Service, Insurance

Who is this presentation for?
Senior Business Executives, Product Managers, Data Scientists/ ML Engineers and High-level Researchers, Product Managers, Data Scientists/ ML Engineers, ML Engineers, Researchers

What you’ll learn:
The audience will learn about what an intelligent application is, how to orchestrate its design, deployment and monitoring in a kubernetes environment. The audience will also learn about the datascientist dilemma and how it may be addressed.

What are the main core message (learning) you want attendees to take away from this talk?
Open source container platforms are a great option to integrate Machine Learning with any application or service by boosting productivity while maintaining a high level of security.

Pre-requisite Knowledge:
Generic SDLC and basic Kubernetes knowledge

What is unique about this speech, from other speeches given on the topic?
Bringing in an enterprise perspective and enterprise ready Kubernetes platform beyond just a proof of concept (POC) although presenting a POC showcase for an end to end intelligent application.

Abstract of Talk:
The recent rise in popularity of containerized workloads demanded better ways to orchestrate and manage these workloads hence the creation of the Kubernetes platform.

When it comes to running intelligent application workloads which contain built-in AI/ML software components, the requirement of a Kubernetes platform as a service extends beyond agility, portability, flexibility and scalability as it is required to also answer to the datascientist’s dilemma: getting started and getting into production. However, as the ML code is only a small part of the entire intelligent application ecosystem, with this workshop we present a showcase for using a Kubernetes platform and a blueprint architecture that proposes an answer to many challenges related to the development, deployment and management of distributed applications.

The user stories we shall focus on in this workshop concerning the developer, data scientist and operations engineer personas are:

- As a data scientist, I want to develop ML models using Jupyter Hub (lab/notebooks) as my preferred research environment.
- As a data scientist, I want my model to be deployed quickly so that it may be used by other applications.
- As a (fullstack) developer, I want to have quick access to resources that support the business logic of my applications, including databases, storage, messaging.
- As a (fullstack) developer, I want an automated build process to support new releases/code updates as soon as they are available in a git repository.
- As an operations engineer, I want an integrated monitoring dashboard to new applications available on the (production) infrastructure.

Can you suggest 2-3 topics for post-discussion?
– DataScientist Kubernetes Platform as a Service
– Automating builds and exposure of ML models inference endpoints

9: 30 AM
-
11:30 PM EST

Graph Neural Network Modeling in Drug Discovery Using PyTorch
Dr. Nasim Abdollahi, Postdoctoral Fellow at University of Toronto, Machine Learning Researcher at Cyclica & Dr. Farnoosh Khodakarami Computer Scientist & ML Researcher, Cyclica

About the Speaker:
Nasim is a Postdoctoral Fellow at University of Toronto and a Machine Learning Researcher Intern at Cyclica, leading a collaborative project between Cyclica, University of Toronto and Vector Institute. She is the vice-chair of Engineering in Medicine and Biology Society of IEEE Toronto section. Nasim obtained her Ph.D. in electrical and computer engineering from University of Manitoba and has M.Sc. and B.Sc. in biomedical engineering. With her passion for developing and applying novel machine learning techniques for improving the quality of health care, she has conducted numerous research projects on enhancing biomedical imaging for breast cancer detection and monitoring. Her current research is focused on graph-based machine learning models that can predict proteins’ biological functions from their 3D atomic structures, with a promise to enhance designing novel medicines. Nasim is an advocate for women in STEM, serves as vice-chair of IEEE Canada Women in Engineering, and was recognized as a “Visionary Emerging Leader”.
Farnoosh Khodakarami is an experienced computer scientist with a demonstrated history of working in the research industry. Skilled in application development with experience in machine learning applications. Strong research professional with a Doctor of Philosophy (Ph.D.) focused in Computer Science. Creative, self-motivated, and committed to working with a team-player attitude, great problem-solving skills, and the ability to quickly grasp new concepts.

Which talk track does this best fit into?
Workshop

Technical level of your talk?
(Technical level: 7/7)

Are there any industries (in particular) that are relevant for this talk?
Hospital & Health Care

What are the main core message (learning) you want attendees to take away from this talk?
Audience will learn about:
– Graph Neural Network (GNN) in drug discovery
– How to build GNN with PyTorch Geometric
– TorchDrug – ML platform for drug discovery
– TorchProtein – a ML library for protein science
– NodeCoder – a graph-based ML framework for predicting proteins’ biological functions

Abstract of Talk:
Graph Neural Networks (GNNs) have been among the most popular neural network architectures, and as graph is a natural representation for protein and molecule, GNNs have shown big sparks in graph-based ML modeling for drug discovery and protein science. Graph-based ML models can help us in identifying the topology of a protein structure from protein sequence, predicting protein’s biological functions from protein structure as well as identifying protein-protein and protein-drug interactions. In this workshop, we will have an introduction on Graph Neural Network (GNN) and its application in drug discovery followed by a code session on PyTorch Geometric, which is a great PyTorch library for building GNN models for structured data. We will then have a code-base session to walk you through two useful tools built with PyTorch Geometric: TorchDrug and NodeCoder.

10: 00 AM
-
11:30 PM EST

Demystifying De-Identification
Patricia Thaine, Co-Founder & CEO, Private AI

About the Speaker:
Patricia Thaine is the Co-Founder & CEO of Private AI, a Microsoft-backed startup, is also a Computer Science PhD Candidate at the University of Toronto (on leave) and a Vector Institute alumna. Her R&D work is focused on privacy-preserving natural language processing, with a focus on applied cryptography and re-identification risk. She also does research on computational methods for lost language decipherment. Patricia is a recipient of the NSERC Postgraduate Scholarship, the RBC Graduate Fellowship, the Beatrice “Trixie” Worsley Graduate Scholarship in Computer Science, and the Ontario Graduate Scholarship. She has ten years of research and software development experience, including at the McGill Language Development Lab, the University of Toronto’s Computational Linguistics Lab, the University of Toronto’s Department of Linguistics, and the Public Health Agency of Canada.

Technical level of your talk?
(Technical Level: 4/7)

What you’ll learn:
Attendees will learn about which privacy enhancing technologies are best for their use case and understand when de-identification is right for them and how not to misuse terminology such as “anonymization”

Abstract of Talk:
Workshop with discussion and demo. The session will begin with an overview of privacy enhancing technologies and then dive into de-identification terminology (de-identification, anonymization, redaction, pseudonymization), how these have been misunderstood, and what to think about when choosing between one of these and other privacy enhancing technologies.
The attendees should bring a sample dataset (preferably made up of unstructured text) and a use case in mind. Each attendee will receive an API key to process a data sample and we will discuss the results. Data can be in languages other than English. Please confirm with organizer that the language is supported first.

12: 00 PM
-
3:00 PM EST

Graph ML – The Next Level of Machine Learning
Jörg Schad, CTO, ArangoDB

About the Speaker:
Jörg Schad is CTO at ArangoDB and enjoys working in intersections of (Graph) databases, Cloud Architectures, and Machine Learning. In a previous life, he has worked on or built machine learning pipelines in healthcare, distributed systems at Mesosphere, and in-memory databases. He received his Ph.D. for research around distributed databases and data analytics. He’s a frequent speaker at meetups, international conferences, and lecture halls.

Which talk track does this best fit into?
Workshop

Technical level of your talk?
(Technical level: 5/7)

Are there any industries (in particular) that are relevant for this talk?
All industries

What are the main core message (learning) you want attendees to take away from this talk?
Machine learning is much more than just building models and the overall pipeline should be considered early on in order to result in actual business impact. Luckily there exist a number of Open-Source projects to help...

Abstract of Talk:
This workshop focuses on why Graphs have become one of the biggest trends in Machine Learning. Graph Machine Learning based on Graph Analytic Algorithms is driving significant improvements in Fraud/Anomaly Detection, Ranking (Page Rank), Recommendation Engines (collaborative filtering), text summarization, and other NLP tasks. We will cover Graph Analytic Algorithms, their applications, and the more novel–but equally exciting–field of Graph Machine Learning, including topics such as Graph Neural Networks, Graph Embeddings, and applications of Graph Machine Learning.

The workshop will be hands-on based on Jupyter notebooks and cover sessions

- Why Graph and Graph Thinking
- Graph Algorithms
- Graph Embeddings
- Graph Neural Networks

12: 00 PM
-
2:00 PM EST

Introduction to NLP & a Step by Step Implementation of a Real World Use Case from TELUS
Mahmudul Hasan, Lead Data Scientist, TELUS Business Marketing

About the Speaker:
With deep expertise in Machine Learning and AI, Mahmudul has over 10 years industry experience of building enterprise level data products to achieve digital transformation, improve customer experience, new revenue opportunity, and cost savings for companies across the globe. He is currently serving as a Lead Data Scientist in TELUS Business Marketing. Mahmudul also designed and developed NLP course content for University of Toronto School of Continuing Studies and also serving as an instructor for the same. Mahmudul holds a Master’s degree in Management Science from University of Waterloo and a Bachelor’s in Computer Science & Engineering.

Which talk track does this best fit into?
Workshop

Technical level of your talk?
(Technical level: 6/7)

Are there any industries (in particular) that are relevant for this talk?
Computer Software, Marketing & Advertising, Telecommunications

Who is this presentation for?
Senior Business Executives, Product Managers, Data Scientists/ ML Engineers and High-level Researchers, Data Scientists/ ML Engineers

What you’ll learn:
The audience will have a real world case study of how unsupervised NLP algorithm can be successfully create values for a business, and some tips and tricks which make this kind of project successful for a data scientist

What are the main core message (learning) you want attendees to take away from this talk?
Audience will see how a business problem is solved leveraging unstructured text data using NLP algorithms along with necessary tips and tricks which makes a unsupervised learning based project financially beneficial for the business.

Pre-requisite Knowledge:
Some basic understanding on Data Science

What is unique about this speech, from other speeches given on the topic?
The audience will get an idea of how unstructured data can be converted to generate financially impactful benefits for business. Also will share some tips on how to make this kind of unsupervised learning based project a successful for a big corporation like TELUS.

Abstract of Talk:
The workshop will be delivered in two part:
Part-1: Brief introduction to NLP concepts and ideas which would include

- Basic definitions and use cases
- Why NLP is a different ball game inside AI/ML (major challenges of processing natural language etc.)
- How those challenges are overcame with ML based approach
- Major workflow of building NLP application.

Part-2: is a detail implementation of a case study with coding details which I have implemented in TELUS. During this part-2, audience will see how a business problem is solved leveraging unstructured text data using NLP algorithms along with necessary tips and tricks which makes a unsupervised learning based project financially successful for the company.

Can you suggest 2-3 topics for post-discussion?
1. What are the challenges of implementing a data science project in business?
2. how can you make your AIML project impactful for the business?

12: 00 PM
-
4:00 PM EST

Beyond the Basics: Data Visualization in Python
*Stefanie Molin, Software Engineer / Data Scientist, **Bloomberg***

About the Speaker:
Stefanie Molin is a software engineer and data scientist at Bloomberg in New York City, where she tackles tough problems in information security, particularly those revolving around data wrangling/visualization, building tools for gathering data, and knowledge sharing. She is also the author of “Hands-On Data Analysis with Pandas,” which is currently in its second edition. She holds a bachelor’s of science degree in operations research from Columbia University’s Fu Foundation School of Engineering and Applied Science. She is currently pursuing a master’s degree in computer science, with a specialization in machine learning, from Georgia Tech. In her free time, she enjoys traveling the world, inventing new recipes, and learning new languages spoken among both people and computers.

Which talk track does this best fit into?
Workshop (1.5-4 hours)

Technical level of your talk?
(Technical level: 4/7)

Who is this presentation for?
Data Scientists/ ML Engineers, ML Engineers, Researchers

What you’ll learn:
A workshop provides the attendees opportunities to ask questions to make sure they are understanding the concepts. Attendees will also have a workshop of curated examples using real-world data rather than the dummy or randomly-generated data nearly everywhere. Each of the visualizations is also created step-by-step, viewing how it changes with each command, which gives attendees a much stronger grasp of the concepts that they can apply elsewhere.

What are the main core message (learning) you want attendees to take away from this talk?
Data visualization is essential for anyone working with data, but sometimes it can be difficult to create impactful visualizations in Python. In this workshop, we will move beyond the plotting basics and explore how to make compelling static, animated, and interactive visualizations.

Pre-requisite Knowledge:
You should have basic knowledge of Python and be comfortable working in Jupyter Notebooks. Check out this notebook for a crash course in Python or work through the official Python tutorial for a more formal introduction. The environment we will use for this workshop comes with JupyterLab, which is pretty intuitive, but be sure to familiarize yourself using notebooks in JupyterLab and additional functionality in JupyterLab. In addition, a basic understanding of pandas will be beneficial, but is not required; reviewing the first section of my pandas workshop will be sufficient.

What is unique about this speech, from other speeches given on the topic?
My teaching style is very different: since the code examples I provide are carefully chosen, it’s easy to see why would take the approach I show, so I make sure that the attendees understand exactly what each line of code is doing to make that happen. I find that this gives the attendees knowledge that they can apply to other problems, rather than just knowing that the code all together has some effect — they get a deeper understanding and can use the concepts like building blocks for their own use cases. Attendees often praise the content in the slides as a detailed reference for later as well.

Abstract of Talk:
The human brain excels at finding patterns in visual representations, which is why data visualizations are essential to any analysis. Done right, they bridge the gap between those analyzing the data and those consuming the analysis. However, learning to create impactful, aesthetically-pleasing visualizations can often be challenging. This session will equip you with the skills to make customized visualizations for your data using Python. While there are many plotting libraries to choose from, the prolific Matplotlib library is always a great place to start. Since various Python data science libraries utilize Matplotlib under the hood, familiarity with Matplotlib itself gives you the flexibility to fine tune the resulting visualizations (e.g., add annotations, animate, etc.). This session will also introduce interactive visualizations using HoloViz, which provides a higher-level plotting API capable of using Matplotlib and Bokeh (a Python library for generating interactive, JavaScript-powered visualizations) under the hood.

Can you suggest 2-3 topics for post-discussion?
Anything relating to the content covered, building data tools, or writing a book/creating workshops

2:30 PM
-
4:00 PM EST

Time Series Anomaly Detection with Machine Learning
*Benjamin Ye, Applied Research Scientists, Georgian & Angeline Yasodhara, Applied Research Scientists, **Georgian***

About the Speaker:
Applied Research Scientists

Which talk track does this best fit into?
Workshop

Technical level of your talk?
(Technical level: 5/7)

What you’ll learn:
Time series anomaly detection methods and applications

Abstract of Talk:
Traditional methods in time series anomaly detection yield good results for relatively simple tasks, but they often fall short when it comes to harder problems of dealing with long-range dependencies, multivariate time series, and subtle contextual anomalies. We introduce a toolkit incorporating classical and novel machine learning techniques (N-BEATS, Transformers, etc.) as well as recent thresholding methods to overcome these challenges. We will discuss their benchmark results against different anomaly types for both univariate and multivariate cases. We will walk through how you can use this simple toolkit and easily incorporate these techniques into your application.

3: 00 PM

4:30 PM EST

NLP for Healthcare: Challenges With Processing and De-Identifying Clinical Notes

Chloe Pou-Prom I Vaakesan Sundrelingam, Data Scientists, **Unity Health Toronto**

About the Speakers:

Chloé Pou-Prom is a data scientist with the Data Science and Advanced Analytics (DSAA) team at Unity Health Toronto. The DSAA team uses high quality healthcare data in innovative ways to catalyze communities of data users and decision makers in making transformative changes that improve patient outcomes and healthcare system efficiency.

Vaakesan Sundrelingam is a data scientist with the GEMINI team at Unity Health Toronto. GEMINI is Canada’s largest hospital data & analytics study, helping physicians, health care teams, and hospitals use data to gain insights into patient care and improve patient outcomes. GEMINI uses machine learning in creative ways to prepare large amounts of data for researchers, as well as in clinical applications such as to detect particularly difficult to measure conditions for quality of care improvement initiatives.

Technical level of your talk?

(Technical Level: 3/7)

What you’ll learn:

1) Why NLP for healthcare is challenging;
2) Why sharing clinical notes across hospitals is difficult; and
3) Some tips and tools to help out with (1) and (2)

Abstract of Talk:

Clinical notes (e.g., admission notes, nurse notes, radiology reports) are rich with information. In this session, we discuss the challenges of working with text data from two different perspectives. First, we provide an overview of the different issues that one can encounter when working with healthcare data, with an emphasis on data processing and cleaning. Then, we focus on the challenges that arise when it comes to sharing data across hospitals, more specifically de-identifying clinical text data. Finally, we provide a demo of pydeid, a Python-based de-identification software that identifies and replaces personal health information (PHI).

3: 30 PM

5:00 PM EST

Distributed Training with PyTorch

Shagun Sodhani, Research Engineer, **Meta AI**

About the Speaker:

Research Engineer at Meta AI, previously at Mila and Adobe Research

Which talk track does this best fit into?

Workshop

Technical level of your talk?

(Technical level: 5/7)

What you’ll learn:

By the end of the session, the attendees would be able to take a simple PyTorch model and scale it to work with dozens of machines. For the straightforward use cases, this will require writing just a few lines of code.

Abstract of Talk:

PyTorch is one of the most popular ML frameworks with the recent releases focusing on enhanced support for distributed training. This talk discusses the different distributed training mechanisms provided by PyTorch. It should be helpful for both practitioners & researchers who want to train larger models and faster.

5: 00 PM

6:30 PM EST

Train your Models Faster by Learning How to Profile and Apply System-Level Optimizations

Akbar Nurlybayev, Co-Founder/VP of Engineering I Xin Li and Yubo Gao, Research Engineers, CentML

About the Speakers:

Akbar is the Co-founder and VP of Engineering at CentML. Previously, Director of Data at KAR Global, \$2 Billion publicly traded company. Yubo and Xin: PhD students at UofT Efficient Computing Lab

Xin Li is a former member of AI Technical Staff at Vector Institute’s AI Engineering team. Working within the vibrant community at Vector Institute, Xin collaborates with Vector researchers and industry partners to make Deep Learning research more accessible in applied settings. Currently, he is working as a Research Engineer at CentML Yubo Gao has recently completed his undergraduate degree at the University of Toronto and have joined the EcoSystem lab as a PhD student, during which he is fortunate to be supervised by Prof. Gennady Pekhimenko.

Which talk track does this best fit into?

Workshop

Technical level of your talk?

(Technical level: 5/7)

What you’ll learn:

This workshop provides unique opportunities for attendees to learn and perform system-level optimizations for deep learning. This topic is often overlooked among ML practitioners due to time and resource constraints. This workshop strives to provide practitioners with some easy-to-use and practical tools to help them understand and optimize their workloads. This workshop also brings a unique perspective on the importance of hardware efficiency when working with Deep Learning models.

Abstract of Talk:

Everybody nowadays train models. Every year the size of the state-of-the-art models grows faster than the hardware becomes cheaper. We observed that many organizations significantly underutilize the available hardware accelerators, i.e. Nvidia GPUs, and as a result, are overpaying for both ML training and inference. In this workshop, our team of world-class ML Systems researchers will share various techniques and tools we use to profile and optimize deep learning models. We will demonstrate how the insights learned from the profiling can be used to discover optimization opportunities that make deep learning models utilize hardware more efficiently. This results in reduced training time, model iteration speed and ultimately lower cost for organizations.

5: 30 PM
-
7:30 PM EST

Iterating on NLP models from R&D to Production
*Denys Linkov, ML Lead, **Voiceflow***

About the Speaker:
Started the ML team at Voiceflow, Kickstarted Rbc’s MLOps journey, Youngest Senior Architect at Rbc. Lead discussion groups and mentorship on MLOps and various blog posts

Technical level of your talk?
(Technical level: 6/7)

Are there any industries (in particular) that are relevant for this talk?
Computer Software, Information Technology & Service, Any startup / large company looking at the R&D process.

Who is this presentation for?
Data Scientists/ ML Engineers, ML Engineers, Researchers

What you’ll learn:
Specific examples and challenges of building NLP products

What are the main core message (learning) you want attendees to take away from this talk?
– How to go from idea to product
– How to iterate on a product
– How to go to production
– How to incorporate customer feedback

What is unique about this speech, from other speeches given on the topic?
Nothing super unique, but many startups of this size rarely share their experience building products and iterating. Many tutorials cover basics and not real business problems.

Abstract of Talk:
Research papers, blogs and products are the culmination of many hours of work, iteration and frustration. However, in these final polished formats, we often gloss over the iteration or creative process on how to get to our desired results.

In this talk, I’ll cover a series of short labs that mirror some of the challenges we’ve faced in building out our NLP models and algorithms. It will be an interactive session with a series of collaborative problem solving, and explanations of what we built and the process we took along the way.
Some of the twists and turns will include:
– Integrating a BERT based model with a message queue system
– Speeding up semantic search through vectorization
– Enabling multi lingual recommendations

Each member of the talk will have access to the code examples and will be encouraged to think beyond the challenges addressed and how they can apply some of our lessons learned to their own work.

Can you suggest 2-3 topics for post-discussion?
BERT based models and embeddings
Deploying models into production
NLP product development

NOVEMBER 29th SESSIONS

9: 59 PM
-
10:04 PM EST

Deploying a Machine Learning Model in under 15 Minutes at Wealthsimple
*Mandy Gu, Engineering Manager, **WealthSimple***

About the Speaker:
Mandy leads the Machine Learning Platform and the Data Platform teams at Wealthsimple. Prior to working on ML-ops and infrastructure, she was a NLP researcher in two separate conversational AI roles and a data scientist building models for the operations and client experience spaces.

Which talk track does this best fit into?
Technical / Research

Technical level of your talk?
(Technical level: 5/7)

Are there any industries (in particular) that are relevant for this talk?
Banking & Financial Services, Computer Software

What are the main core message (learning) you want attendees to takeaway from this talk?
The "power of platform" and how investing in strong ML foundations can accelerate model development

What are some of the infrastructure you plan to discuss?
How we built the first iteration of our ML Platform, from idea to operational microservices for model training, deployment and prediction serving (and the process of migrating existing workflows onto the platform)

What are some of the languages you plan to discuss?
Primarily Python

What kind of Dev ops tools you plan to discuss? Open source?
MLflow, FastAPI, NVIDIA Triton Inference Server

On a scale of 1-10 how mature is this applied AI application you plan to discuss?
10

Who is this presentation for?
Data Scientists/ ML Engineers

Pre-requisite knowledge:
Machine learning model development lifecycle, high level concepts of what it means to deploy a model to production

What you'll learn:
We break down the process of how we got from no ML Platform to a fully operational ML Platform in under a quarter. We learned a lot through this experience, and a lot of these learnings aren't easily discoverable through a Google search :)

What is unique about this speech, from other speeches given on the topic?

The speech is unique to Wealthsimple's experience with ML infrastructure, tooling and deployment. When it came to designing our platform, we drew inspiration from the Rails Doctrine to accelerate ML by convention over configuration and putting the developer experience on a pedestal. Our goal is to build something revolutionary, just like how Ruby on Rails transformed web development 20 years ago.

Abstract of Talk:
From powering money movement to fraud detection, machine learning models are critical to Wealthsimple's core business process. This year, we built our next generation Machine Learning platform with a simple goal in mind: deploy new ML models within HOURS. This is how we scoped, designed and built our platform in just under 3 months.

10: 06 PM
-
10:11 PM EST

What To Look For In Your Next ML Pipeline
*Bhavani Rao, Technical Product Marketing Manager, **Pachyderm***

About the Speaker:
Bhavani Rao is a Technical Product Marketing Manager, responsible for product messaging and positioning at Pachyderm, a leader in data pipelining and MLOps. He has a diverse background, working with customers in Data Ops, DevOps, CI/CD, relational and NoSQL databases. A recent convert to the potential of AI/ML, Bhavani is passionate about technology and how it can be leveraged to solve customer problems. Throughout his career, Bhavani has promoted these learnings and best practices at numerous industry gatherings. He has a B.S. degree in Operations Research from Indiana University and an MBA from Columbia University.

Abstract of Talk:
MLOps is not the same as DevOps. Iteration is a common theme to both methodologies but the requirements are different. Your pipelines need to version the code AND the data for easy reproducibility and rollback. Given the enormous size of datasets, data pipelines need to scale to petabytes, automatically trigger and process only the new data, rather than executing a complete run every time. Join us for this lightning talk as we discuss: what are data pipelines and how to leverage pipelines to quickly converge on a ML model.

10: 13 PM
-
10:18 PM EST

Retraining won't fix your model (always)
*Oren Razon, Co-founder & CEO, **Superwise***

About the Speaker:
Oren is the co-founder and CEO of Superwise, the leading platform for model observability. With over 15 years of experience leading the development, deployment, and scaling of ML products, Oren is an expert ML practitioner specializing in MLOps tools and practices. Previously, Oren managed machine learning activities at Intel's ML center and operated a machine learning boutique consulting agency helping leading tech companies such as Sisense, Gong, AT&T, and others, to build their machine learning-based products and infrastructure.

Abstract of Talk:
It's practically dogma today that a model's best day in production will be its first day in production. Over time model performance degrades, and there are many variables that can cause decay, from real-world behavior changes to data drifts. When models misbehave, we often turn to retraining to fix the problem, but retraining is not always the best or only solution out there. In this session we'll take a crash intro in alternative techniques.

10: 18 PM
-
10:23 PM EST

7 Questions for Data Scientists
*Geoffrey Hunter, Lead Data Scientist, **Spothero***

About the Speaker:
Geoffrey is passionate about forming end-to-end, product-focused Data Science teams that deliver high impact results. After his post doc, he was a Data Science consultant at different companies and then moved onto leading Data Science teams. He acts to contextualize Data Science opportunities for senior leadership and then mobilizes and mentors the data science teams to focus on understanding and solving problems.

Abstract of Talk:
Geoffrey will share the 7 Questions I ask on a daily basis to rapidly qualify and frame new Data Science problems. This framework can be applied to understand new opportunities as well as to existing problems to help eliminate noise and focus one's efforts.

10: 55 PM
-
11:30 PM EST

Saving Lives with ML: Applications and Learnings
Muhammad Mamdani, Unity Health Toronto – VP: Data Science and Advanced Analytics; Director: Temerty Centre for Artificial Intelligence Research and Education in Medicine of the University of Toronto; Professor – University of Toronto

About the Speaker:
Dr. Mamdani is Vice President of Data Science and Advanced Analytics at Unity Health Toronto and Director of the University of Toronto Temerty Faculty of Medicine Centre for Artificial Intelligence Research and Education in Medicine (T-CAIREM). Dr. Mamdani’s team bridges advanced analytics including machine learning with clinical and management decision making to improve patient outcomes and hospital efficiency.

Dr. Mamdani is also Professor in the Department of Medicine of the Temerty Faculty of Medicine, the Leslie Dan Faculty of Pharmacy, and the Institute of Health Policy, Management and Evaluation of the Dalla Lana Faculty of Public Health. He is also adjunct Senior Scientist at the Institute for Clinical Evaluative Sciences (ICES) and a Faculty Affiliate of the Vector Institute. In 2010, Dr. Mamdani was named among Canada’s Top 40 under 40. He has published over 500 studies in peer-reviewed medical journals.

Dr. Mamdani obtained a Doctor of Pharmacy degree (PharmD) from the University of Michigan (Ann Arbor) and completed a fellowship in pharmacoconomics and outcomes research at the Detroit Medical Center. During his fellowship, Dr. Mamdani obtained a Master of Arts degree in Economics from Wayne State University in Detroit, Michigan with a concentration in econometric theory. He then completed a Master of Public Health degree from Harvard University in 1998 with a concentration in quantitative methods.

Which talk track does this best fit into?
Case Study

Technical level of your talk?
(Technical level: 3/7)

Are there any industries (in particular) that are relevant for this talk?
Hospital & Health Care

Who is this presentation for?
The successful application of ML in healthcare is multifaceted and highly dependent on end-user engagement. Innovative public-private partnerships are needed to spread ML applications globally. Multidisciplinary, collaborative efforts will fuel innovations in the development and application of ML in healthcare.

Abstract of Talk:
Machine learning (ML) has transformed numerous industries but its application in healthcare has been limited. ML applications are expected to permeate healthcare in the near future with a recent explosion in academic and commercial activity. The application of ML in healthcare, however, is complicated by a variety of factors including the significant variability in needs, healthcare settings and patients served in these settings, workflows, and available resources. This talk will present a case study of Unity Health Toronto and its journey in developing and deploying numerous ML solutions into clinical practice, including bridging public and private sector partnerships to spread innovations internationally. The talk will also present a novel Canadian academic centre dedicated to artificial intelligence (AI) in medicine – the Temerty Centre for Artificial Intelligence Research and Education in Medicine (T-CAIREM) at the University of Toronto.

10: 55 PM
-
11:30 PM EST

Scaling Advanced Analytics in the Worst Crisis in the Industry Area
Nicolás Venegas Oliva, Technical Lead of Advanced Analytics | Cristobal Guzman, Staff Data Scientist, **LATAM Airlines**

About the Speaker:
2 years of experience in backend development, 2+ years in data processing and the last 3+ years as Advanced Analytics technical leader at LATAM Airlines. During this time the team has grown from 9 to 48 highly trained professionals. It has also become the team with the highest impact generation within the company and a reference in the region in terms of MLOps and measured business impact through data products.

Which talk track does this best fit into?
Case Study

Technical level of your talk?
(Technical level: 5/7)

What you'll learn:
Scaling of MLOps teams, impact measurement, selection and training of highly technical teams.

Are there any industries (in particular) that are relevant for this talk?
Computer Software, Information Technology & Service

What are the main core message (learning) you want attendees to takeaway from this talk?
Scaling of MLOps teams, impact measurement, selection and training of highly technical teams.

Abstract of Talk:
For data science teams looking to create real business value with AI – MLOps is not something that’s ‘nice to have’ – It’s a MUST HAVE. To make MLOps work for your organization, you need to have the right tools combined with the right skillset across the different roles, and a unified process. For LATAM Airlines Group, being faced with the worst airline industry crisis following the COVID-19 pandemic, MLOps was imperative. We set off to create a cross-company MLOps strategy and implement it across dozens of use cases. In this talk, we will share our MLOps strategy, provide tips for success, pitfalls to avoid based on our own data science journey and dive into two of our use cases.

10: 55 PM
-
11:30 PM EST

Where's the Road? The Challenge of Autonomous Driving Perception in Winter
Steven Waslander Professor, *Institute for Aerospace Studies* / Director, *Toronto Robotics and AI Laboratory*, *University of Toronto*

About the Speaker:
Prof. Steven Waslander is a leading authority on autonomous aerial and ground vehicles, including multirotor drones and autonomous driving vehicles. He received his B.Sc.E.in 1998 from Queen's University, his M.S. in 2002 and his Ph.D. in 2007, both from Stanford University in Aeronautics and Astronautics. He founded and directed the Waterloo Autonomous Vehicle Laboratory (WAVELab) in 2008-2018, and the Toronto Robotics and Artificial Intelligence Laboratory (TRAILab) at the University of Toronto from 2018 onward.

Prof. Waslander’s work on autonomous vehicles has resulted in the Autonomoose, the first autonomous vehicle created at a Canadian University to drive on public roads. His insights into autonomous driving have been featured in the Globe and Mail, Toronto Star, National Post, and the Rick Mercer Report. He has over 160 publications and host the Self-Driving Car Specialization on Coursera, which has accumulated over 150,000 learners worldwide since 2019.

Which talk track does this best fit into?
Technical

Technical level of your talk?
(Technical level: 6/7)

What you'll learn:
That winter is harder than clear weather, but that we can still build safe self-driving cars for any weather condition, if we take the time to work through the added challenges.

Abstract of Talk:
Autonomous driving solutions are steadily progressing toward real-world deployments, but most companies are focused on driving in clear weather days in benign climates. Our work on exposing the challenges of Canadian winters for perception tasks has led to the Canadian Adverse Driving Conditions Dataset, and to multiple advances in all-weather autonomy that set the stage for more complete dominion of robotics systems over sensor degradation due to precipitation and accumulation. In this talk, I'll highlight some of the worst problems that arise for autonomous systems in Canada, and lay out our plans for WinTOR, a new University of Toronto research program aimed at helping self-driving vehicles extend their range to our roadways year round.

11: 35 PM
-
12:05 PM EST

From Silo to Collaboration - Building Tooling to Support Distributed ML Teams at Twitch
Shiming Ren, Sr. Engineering Manager – Safety, MLOps and Infrastructure & Chen Liu, Twitch Sr. Engineering Manager on Personalization and ML Infra, Amazon/Twitch

About the Speaker:
I Shiming worked as a Software Engineer Manager at Twitch about MLOps and Tooling in Safety team. I spoke at Meta’s At Scale about Scaling ML Workflows for Real-Time Moderation Challenges at Twitch, I also spoke at TwitchCon about Integrating Data into Twitch at Scale. I worked in engineering leadership role for 5 years and our team made several company wide MLOps tooling such as orchstration and feature store.
Chen is currently supporting teams working on personalization and ML infrastructures at Twitch. He is passionate about building scalable ML products and democratizing ML in the organization.

Which talk track does this best fit into?
Technical / Research

Technical level of your talk?
(Technical level: 4/7)

Are there any industries (in particular) that are relevant for this talk?
Computer Software, Information Technology & Service

Who is this presentation for?
Senior Business Executives, Product Managers, Data Scientists/ ML Engineers and High-level Researchers, Data Scientists/ ML Engineers, ML Engineers

What you’ll learn:
Twitch’s strategy of scaling our ML infra and MLOps tooling has never been discussed online. And we aim to help audience figure out the best strategy to utilize ML tooling for enhancing collaborations between ML teams and boost scientists self-service / efficiency. This is a good lesson if companies are seeking to start MLOps from stratch.

On a scale of 1-10 how mature is this applied AI application you plan to discuss?
7/10

Pre-requisite Knowledge:
Feature store, Orchstration, Large Scale Data Handling

What kind of DevOps tools you plan to discuss? Open source?
N/A Our tools are all in house

What are some of the languages you plan to discuss?
Python, Golang

What are some of the infrastructures you plan to discuss?
Feature Store, ML Orchstration, Realtime Inference, Distributed ML team collaborations

What is unique about this speech, from other speeches given on the topic?
We aim to use examples how Twitch build in house feature store, realtime inference and orchstration system to demonstrate from technology perspective about MLOps collaborations in a company. This is more like a hybrid tech and management talk which will benefit both engineer and leadership groups.

Abstract of Talk:
In this talk, we will cover Twitch’s current ML team structure and its challenges of it. Then we dive deep into some solutions we have built to support ML development at Twitch, including what they are and how they will benefit the situation. We close with a discussion of Twitch’s distributed ML team style and how we collaborate using Conductor as an example.

ML has been playing a more and more important role in Twitch’s products (e.g. Recommendation, Safety). In order to allow products to iterate fast, we keep ML practitioners in the product teams and empower the teams to work independently. Undoubtedly, there are common challenges in ML development regardless of product areas. So we are striving to develop tooling and infrastructures for general ML development in order to reduce duplicate work across ML teams. We will dive into those efforts we made in this presentation. For example, Twitch machine learning feature store is developed to have a single control plane serving as feature registry but facilitates distributed feature ownership (e.g. storage, pipelines). Conductor, a in-house ML orchestration system, promotes best practices in pipeline management with templated process control flow and distributed infrastructure management. Meanwhile, we are promoting collaborative ML culture among Twitch engineering teams. It is similar to community-owned open source projects where teams share the same interests and encourage cross team contribution and development.
Can you suggest 2-3 topics for post-discussion?
Manage ML teams collaboration in a distributed manner; ML tooling development from 0 to 10; Implementation details for feature store and ML orchestration system.

11: 35 PM
-
12:05 PM EST

Saga: Continuous Construction and Serving of Large Scale Knowledge Graphs
Ihab Ilyas, Professor in the Cheriton School of Computer Science and the NSERC-Thomson Reuters Research Chair on Data quality at the University of Waterloo

About the Speaker:
Ihab Ilyas is a professor in the Cheriton School of Computer Science and the NSERC-Thomson Reuters Research Chair on data quality at the University of Waterloo. He is currently on leave at Apple to lead the Apple Knowledge Platform team. His main research focuses on the areas of Data Science and data management, with special interest in data quality and integration, managing uncertain data, machine learning for data curation, and information extraction. Ihab is a co-founder of Tamr, a startup focusing on large-scale data integration, and the co-founder of inductiv (acquired by Apple), a Waterloo-based startup on using AI for structured data cleaning. He is an ACM Fellow and IEEE Fellow, a recipient of the Ontario Early Researcher Award, a Cheriton Faculty Fellowship, an NSERC Discovery Accelerator Award, and a Google Faculty Award. Ihab was an elected member of the VLDB Endowment board of trustees (2016-2021), elected SIGMOD vice chair (2016-2021), an associate editor of the ACM Transactions of Database Systems (2014-2020), and an associate editor of Foundations of Database Systems. He holds a PhD in computer science from Purdue University, West Lafayette.

Which talk track does this best fit into?
Technical / Research

Technical level of your talk?
(Technical level: 5/7)

Are there any industries (in particular) that are relevant for this talk?
Information Technology & Service

What are the main core message (learning) you want attendees to take away from this talk?
Complexity of building large scale knowledge graphs.

Abstract of Talk:
In this talk I present Saga, an end-to-end platform for incremental and continuous construction of large scale knowledge graphs we built at Apple. Saga demonstrates the complexity of building such platform in industrial settings with strong consistency, latency, and coverage requirements. In the talk, I will discuss challenges around the following: building source adapters for ingesting heterogenous data sources; building entity linking and fusion pipelines for constructing coherent knowledge graphs that adhere to a common controlled vocabulary; updating the knowledge graphs with real-time streams; and finally, exposing the constructed knowledge via a variety of services. Graph services include: low-latency query answering; graph analytics; ML-biased entity disambiguation and semantic annotation; and other graph-embedding services to power multiple downstream applications. Saga is used in production at large scale to power a variety of user-facing knowledge features.

1: 15 PM
-
1:45 PM EST

Declarative Machine Learning Systems: Ludwig & Predibase

Predibase - Piero Molino Talk Title

1: 15 PM
-
1:45 PM EST

The Application of Mobile Location Data for Vending Machine Site Selection and Revenue Optimization
Nikita Medvedev, Director of Advanced Analytics & Winston Li, Founder, Coca Cola & Arima

About the Speaker:
Winston is the founder of Arima, a Canadian based startup that provides consumer data to its users. Our flagship product, the Synthetic Society, is a privacy-by-design, individual level database that mirrors the real society. Built using trusted sources like census, market research, mobility and purchase patterns, it contains 10k+ attributes across North America and enables advanced modelling at the most granular level.
Prior to founding Arima, Winston was the Director of Data Science at PwC and Omnicom. Winston is also a part-time faculty member at Northeastern University Toronto and sits on the advisory board of the Master of Analytics program.

Nikita is the Director of Advanced Analytics at Coca-Cola Canada Bottling Limited. Together with his team he is transforming terabytes of business operations data into actionable insights to drive growth and innovate in the Consumer Packaged Goods industry. He loves finding novel solutions to old problems and is obsessed with driving real lasting change through better use of data.
Nikita has over 10 years of experience in the Retail and Consumer Packaged Goods industries, working for companies like Loblaw and Sears. He is also an alumnus of the Master of Management Analytics program from Queen’s University, and holds a Bachelor of Finance & Economics degree from University of Toronto.

Which talk track does this best fit into?
Case Study

Technical level of your talk?
(Technical level: 4/7)

Are there any industries (in particular) that are relevant for this talk?
Food & Beverages, Information Technology & Service, Marketing & Advertising

What are the main core message (learning) you want attendees to take away from this talk?
Mobility data as an alternative data source for consumer related analytics and its recency and granularity and really drive measurable business outcomes.

Abstract of Talk:
In this presentation, we present an innovative approach to utilizing mobility data to optimize the placement of vending machines in Canada. Coca-Cola has more than 10k vending machines in various locations and their ROI heavily depends on the amount of foot traffic next to them as well as who those people are. For this use case, we’ll be concentrating on using the super detailed mobility data to understand the difference between our best machines and worst at scale, and optimizing their location based on the mobility data to increase the ROI. In addition to the practical and business application, we’ll also be able to share the algorithms used and the tech stack with the audience.

1: 15 PM
-
1:45 PM EST

The Risks of Excluding the Disengaged From your Dataset
Danielle Goldfarb, Vice President and General Manager of Global Affairs, Economics and Public Policy, RIWI (Real-Time Interactive Worldwide Intelligence)

About the Speaker:
Opening speaker, Tedx Toronto: The Smartest Way to Predict the Future; Developing a U of Toronto course on new data tools for global affairs / public policy.

Which talk track does this best fit into?
Business Strategy

Technical level of your talk?
(Technical level: 1/7)

What you’ll learn:
There is a huge emphasis on tools to analyze big data but we need to spend at least as much time on the quality of the underlying dataset. If, for example, we exclude disengaged populations, as some common methods do, we risk getting things wrong, whether understanding a market’s potential, predicting an election result, or anticipating the trajectory of the pandemic, the war, or the economy.

Abstract of Talk:
Analysts use machine learning tools to analyze the vast amounts of data now available. Big data is appealing, but even the best tools and techniques will give you noise if the underlying data aren’t robust and inclusive. To reliably predict elections, correctly anticipate consumer demand or to accurately predict the trajectory of a pandemic, leaders need to ask themselves who is (and isn’t) included in their dataset.

1: 50 PM
-
2:20 PM EST

Real-time Machine Learning: Architecture and Challenges
Chip Huyen, CEO, Claypot AI

About the Speaker:
Chip Huyen is a co-founder of Claypot AI, a platform for real-time machine learning. Previously, she was with Snorkel AI and NVIDIA. She teaches CS 329S: Machine Learning Systems Design at Stanford. She’s the author of Designing Machine Learning Systems, an Amazon bestseller in AI. She has also written four bestselling Vietnamese books.

Which talk track does this best fit into?
Technical / Research

Technical level of your talk?
(Technical level: 5/7)

Are there any industries (in particular) that are relevant for this talk?
Banking & Financial Services, Computer Software, Information Technology & Service, Insurance, Marketing & Advertising

What are the main core message (learning) you want attendees to take away from this talk?
Fresh data beats stale data for machine learning applications

Abstract of Talk:
Fresh data beats stale data for machine learning applications. This talk discusses the value of fresh data as well as different types of architecture and challenges of online prediction.

1: 50 PM
-
2:20 PM EST

Navigating the Tradeoff Between Privacy and Fairness in ML
Jesse Cresswell, Senior Machine Learning Scientist, Layer 6 AI

About the Speaker:
Jesse is a Senior Machine Learning Scientist at Layer 6 AI within TD, and is the Team Lead for Credit Risk. His applied work centers on building machine learning models in high risk and highly regulated domains. Jesse leads research on privacy enhancing technologies for machine learning including topics of Federated Learning and Differential Privacy.

Which talk track does this best fit into?
Technical

Technical level of your talk?
(Technical Level: 5/7)

What you’ll learn:
Applying privacy enhancing technologies can increase bias and unfairness in ML models. Practitioners need to consider the intersection of these important ethical ideas.

Abstract of Talk:
As machine learning becomes more widespread throughout society, aspects including data privacy and fairness must be carefully considered, and are crucial for deployment in highly regulated industries. Unfortunately, the application of privacy enhancing technologies often worsens unfair tendencies in models. In this talk we address the intersection of privacy and fairness in machine learning, and offer research-based solutions for navigating the tradeoffs.

2: 55 PM
-
3:25 PM EST

Optimal Beer Pricing: An Optimization Layer for Price Elasticities
Eric Hart, Staff Data Scientist at Anheuser-Busch

About the Speaker:
Eric is a Staff Data Scientist with more than 7 years of experience working at Altair Engineering and Anheuser-Busch. He has a PhD in probability from the University of Toronto, and a masters degree in Applied Math and an undergraduate degree in Engineering from Queen’s university. He’s also a world champion Blokus player.

Which talk track does this best fit into?
Case Study

Technical level of your talk?
(Technical level: 2 /7)

Are there any industries (in particular) that are relevant for this talk?
Banking & Financial Services, Food & Beverages, Marketing & Advertising

Who is this presentation for?
Senior Business Executives, Product Managers, Data Scientists/ ML Engineers and High-level Researchers

What you’ll learn:
Putting a mathematical optimization layer on top of predictive models is still a mostly unused tool in the ML space. It’s very difficult to learn about that from existing resources.

What are the main core message (learning) you want attendees to take away from this talk?
How to add an optimization layer to ml models.

Pre-requisite Knowledge:
Not a lot. We’ll briefly discuss what price-elasticities and mathematical optimization are, but having heard those terms before (with a basic understanding) would help.

What is unique about this speech, from other speeches given on the topic?
I would argue the whole topic is fairly unique (optimization layers for predictive models are not widely used or discussed). In addition, the specifics of trying to work around the realities of the beer industry (especially varying laws about beer pricing across different geographies) add an extra layer of complexity to this already deep problem.

Abstract of Talk:
At Anheuser-Busch, we’re obsessed with price elasticities. When the price of beer changes, how will that affect the volume of beer that we sell? These questions (yes, this is more than one question) have implications all over the business, from price setting to procurement to financial planning. We’ve worked hard to make sure our answers to these questions are as data driven as possible. But once we have a model to produce (and predict) these elasticities, how do we make business decisions based on that? And how do we make sure those business decisions are also as data driven as possible?

In this talk we’ll discuss an optimal pricing layer for beer elasticities. We’ll cover how to use mathematical optimization to make specific price change suggestions at a variety of granularities to help achieve specific business objectives. We’ll consider what objective we actually want to optimize (Profit? Revenue? Market Share?) and see how to use constraints to help smooth the trade-off between these objectives. Finally, we’ll investigate how to ensure our price suggestions stay within the regions where the underlying elasticities models make sense. Ever wanted to see a real-world example of levelling up your analytics from predictive- to prescriptive-, and do so in the context of price setting (or beer drinking)? Now’s your chance!

Can you suggest 2-3 topics for post-discussion?
Optimization Layers. Price Elasticities.

2: 55 PM
-
3:25 PM EST

Outracing Champion Gran Turismo Drivers With Deep Reinforcement Learning
*Varun Raj Kompella, Senior Research Scientist, **Sony AI***

About the Speaker:

Varun Kompella is currently a senior research scientist at Sony AI. He earned his master’s of science degree in informatics with a specialization in graphics, vision and robotics from Institut Nationale Polytechnique de Grenoble (INRIA Grenoble), and a Ph.D degree from Università della Svizzera Italiana (IDSIA Lugano), Switzerland, working with Prof. Juergen Schmidhuber. In his thesis work he developed algorithms that use the slowness principle for driving exploration in reinforcement learning agents. After completing his Ph.D., he worked as a postdoctoral researcher at the Institute for Neural Computation (INI), Germany. His research contributions led to several patents, publications in peer-reviewed journals and conference proceedings.

Which talk track does this best fit into?

Technical / Research

Technical level of your talk?

(Technical Level: 7/7)

What you’ll learn:

We demonstrate the possibilities and challenges of using deep RL techniques to control complex dynamical systems in domains such as Gran Turismo where agents must respect imprecisely defined human norms.

Abstract of Talk:

Many potential applications of artificial intelligence involve making real-time decisions in physical systems while interacting with humans. Automobile racing represents an extreme example of these conditions; drivers must execute complex tactical manoeuvres to pass or block opponents while operating their vehicles at their traction limits. Racing simulations, such as the PlayStation game Gran Turismo, faithfully reproduce the non-linear control challenges of real race cars while also encapsulating the complex multi-agent interactions. Here we describe how we trained agents for Gran Turismo that can compete with the world’s best e-sports drivers. We combine state-of-the-art, model-free, deep reinforcement learning algorithms with mixed-scenario training to learn an integrated control policy that combines exceptional speed with impressive tactics.

In addition, we construct a reward function that enables the agent to be competitive while adhering to racing’s important, but under-specified, sportsmanship rules. We demonstrate the capabilities of our agent, Gran Turismo Sophy, by winning a head-to-head competition against four of the world’s best Gran Turismo drivers. By describing how we trained championship-level racers, we demonstrate the possibilities and challenges of using these techniques to control complex dynamical systems in domains where agents must respect imprecisely defined human norms.

2: 55 PM
-
3:25 PM EST

Bias and Relevance in Application of ML in Healthcare

*Shazia Akbar, Lead Machine Learning Engineer, **Altis Labs***

*Ali Madani, Director of Machine Learning, **Cyclica***

*Santosh Hariharan, Principal Scientist, **Pfizer***

*Shiva Amiri, VP, Head of AI and Data Intelligence, **Pivotal Life Sciences***

*Javier Diaz-Mejia, Head of Data Science, **Phenomic AI***

About the Speaker:

Dr. Shazia Akbar is the lead machine learning engineer at Altis Labs, a Toronto-based startup which leverages deep learning technologies to gain prognostic insight from medical imaging data. Since joining Altis Labs, Shazia has designed and developed artificial intelligent systems which ingest millions of imaging data to predict patient outcomes. Some of the applications she has developed to date include a fully automated model to quantify mortality risk in early stage lung cancer patients, and an x-ray model which determines in-patient admission risk of hospital patients diagnosed with community acquired pneumonia.

Shazia gained her PhD from the University of Dundee, UK, after which she joined the department of Radiology at New York University, US. In 2018, Shazia completed her postdoctoral fellowship at Sunnybrook Research Institute and the Vector Institute, designing novel deep learning algorithms for digital pathology. Her research interests include explainable AI, weakly supervised learning and applications of AI in healthcare.

Ali Madani leads ML technology development at Cyclica, a leading Canadian biotechnology company focused on AI based drug discovery. He is also editor of special Topic Artificial Intelligence in Cancer Diagnosis and Therapy at MDPI and work as an AI educator with companies like WeCloudData. Ali is a PhD graduate of University of Toronto; an alumni of University of Waterloo School of Engineering; and, attained a master of mathematics from the University of Waterloo. He is an active member of the machine learning community in Toronto and speaks in world and Canada wide conferences, webinars and workshops about technology development, machine learning, drug discovery and cancer therapeutics. He has also published more than 20 scientific articles in high impact factor journals on these subjects.

Santosh Hariharan is a Principal Scientist at Pfizer, committed to curing disease and improving patient lives. He enjoys solving complex biological problems using simple blocks with a motto of "Seeing is Believing". He develops and analyzes complex biology by looking at individual cells, evaluating their response to drugs/genetic perturbation and developing predictive models using AI and machine learning (Phenotyping).

Shiva Amiri is the VP, Head of AI and Data Intelligence at Pivotal Life Sciences, working at the intersection of computing and biology with experience in large-scale, multi-stakeholder technology development in data science and biology with a focus on computational biology, bioinformatics, machine learning, and big data systems. Shiva is a team builder and entrepreneurial in cutting edge computational methods in biology, digital health and medical research with a track record in big data/data science and program execution and strategy.

Javier is a data scientist with 15+ years of experience in projects aiming to solve problems of relevance for human health. He has experience in the academic, non-profit and industry sectors in Mexico, USA and Canada. Javier's role involves identifying organization data science needs, building teams to implement solutions addressing those needs, and serving as a bridge between technical and executive stakeholders. Javier made his PhD studies in Mexico, postdoctoral training in Toronto and currently, works as Head of Data Science at Phenomic AI, a biotech startup developing machine learning tools to fight cancer.

Which talk track does this best fit into?

Technical / Research

Technical level of your talk?

(Technical Level: 4/7)

Are there any industries (in particular) that are relevant for this talk?

Hospital & Health Care

What you’ll learn:

Examples of application of ML in healthcare; Impact of ML in healthcare technologies on patients; Potential biases in ML in Healthcare technologies

Abstract of Talk:

Machine Learning (ML) technologies learn how to accomplish tasks and identify patterns from available data. This data-based learning has been critical in helping the healthcare institute, biotechnology and pharmaceutical companies in developing new technologies to improve processes like disease diagnosis from radiological and pathological images and drug design. In this panel, we will discuss how companies design ML technologies that are not only in-line with their business model but eventually will impact patients and healthcare systems. We will also discuss the technological and sociological biases that need to be taken into account in the design of such technologies.

3: 30 PM
-
4:00 PM EST

Building a Fully Automated ML Platform Using Kubeflow and Declarative Approach to Development of End-to-End ML Pipelines
*Valerii Podymov, Lead Data Scientist | Roshan Issac, Machine Learning Engineer | Vlad Ryzhkov, Senior Data Engineer | Joey Zhou, Senior Data Engineer at **FreshBooks***

About the Speaker:
Valerii joined FreshBooks a year ago to lead and grow a team of Data Scientists and Machine Learning Engineers. He has an experience in multiple industries ranging from Electronics to Clean Tech and has contributed to the development of innovative solutions for a variety of brands such as LG Electronics, Panasonic, Samsung, Toyota, Scotiabank, Cineplex. He has a University Degree in Telecom Engineering and PhD in Automated Control Systems. Author of 20 patented inventions in Signal Processing, Electronics and Computing.

Roshan Issac, Machine Learning Engineer at FreshBooks
Roshan works as a Machine Learning Engineer at FreshBooks where he is building ML Platform on Vertex AI and bringing MLOps best practices to the organization. He was previously at the same role with Cineplex. He has a Bachelor Degree in Computer Science and Engineering and hold graduate certificates in AI & Project Management. Overall he has 8+ years of experience in Machine Learning, Data Analytics and CRM software working in different startups and companies in Canada and India. He published papers in IEEE conferences and was a speaker at Libre Software Meeting (LSM), France.

Vlad Ryzhkov, Senior Data Engineer at FreshBooks
Vlad joined FreshBooks a year ago with extensive Data Engineering background and he works on building ML Platform bringing best practices in large-scale data processing to the company. He has a PhD in System Analysis, Management and Information Processing. Overall, his 15+ years of software development experience comprises such areas as financial systems, e-commerce, e-sport and airlines in Canada and overseas.

Joey Zhou, Senior Data Engineer at FreshBooks
Joey joined FreshBooks three months ago and works on the continuous monitoring framework for the ML team. Before, he had an experience in the tech industry, ranging from social-dating to e-commerce, in multiple roles such as Data Scientist and Machine Learning Engineer. He built a recommender systems for one of the largest e-commerce platforms in China. With hands-on experience in building and productionizing ML models, he is ready to pursue his passion for MLOps at FreshBooks.

Which talk track does this best fit into?
Technical / Research

Technical level of your talk?
(Technical level: 6/7)

Are there any industries (in particular) that are relevant for this talk?
Banking & Financial Services, Computer Software

Who is this presentation for?
Data Scientists/ ML Engineers, ML Engineers

What you'll learn:
How we tackled existing challenges with Kubeflow pipelines changing the imperative approach to the declarative.

What are the main core message (learning) you want attendees to take away from this talk?
A journey to higher levels of MLOps maturity is unique for any company and has no recipes due to experimental nature of MLOps. Many insights and ideas in this area are the results of investments by big names (Google, Microsoft, Amazon) and knowledge sharing between smaller companies like us working on similar problems. We are grateful for this opportunity to contribute to the ecosystem so that others can learn from us.

On a scale of 1-10 how mature is this applied AI application you plan to discuss?
9/10

Pre-requisite Knowledge:
Machine Learning Lifecycle

What kind of DevOps tools you plan to discuss? Open source?
GitHub Actions, Kubeflow

What are some of the languages you plan to discuss?
Python, SQL

What are some of the infrastructures you plan to discuss?
BigQuery, Airflow, Vertex AI, containers

What is unique about this speech, from other speeches given on the topic?
Managing MLOps is highly immature topic with lack or absence of commonly accepted best practice, so the experience of any company in growing over MLOps maturity levels is always unique.

Abstract of Talk:
This talk is about our journey at FreshBooks from mostly manual processes in productionizing of our ML models to the highest levels of maturity in MLOps. First, we briefly go over a list of challenges we faced when working on the ML platform as a hybrid team of Data Scientists, ML Engineers and Data Ops Engineers. And then we provide more detailed overview of our end-to-end Kubeflow pipelines and a declarative MLOps framework that has been designed to speed up, simplify and improve the reliability of ML pipelines at each stage from development to production. Lessons learned and what's next will be provided at the end of the talk.

Can you suggest 2-3 topics for post-discussion?
ML Ops, ML Model Governance

3: 30 PM
-
4:00 PM EST

PyScript for Data Science
*Sophia Yang, Senior Data Scientist, **Anaconda***

About the Speaker:
Sophia Yang is a Senior Data Scientist at Anaconda, Inc., where she uses data science to facilitate decision-making for various departments across the company. She volunteers as a Project Incubator at NumFOCUS to help Open Source Scientific projects grow. She is also the author of multiple Python open-source libraries such as condastats, cranlogs, PyPowerUp, intake-stripe, and intake-salesforce. She holds an M.S. in Statistics and Ph.D. in Educational Psychology from The University of Texas at Austin.

Which talk track does this best fit into?
Technical / Research

Technical level of your talk?
(Technical level: 3/7)

Are there any industries (in particular) that are relevant for this talk?
Computer Software, Information Technology & Service

Who is this presentation for?
Senior Business Executives, Product Managers, Data Scientists/ ML Engineers and High-level Researchers, Product Managers, Data Scientists/ ML Engineers

What you’ll learn:
PyScript was announced early this year. There is not much tutorial and content online yet.

Pre-requisite Knowledge:
Knowledge of Python is recommended

What is unique about this speech, from other speeches given on the topic?
I will introduce PyScript from a data science perspective.

Abstract of Talk:
Are you a data scientist or a developer who mostly uses Python? Are you jealous of developers who write Javascript code and build fancy websites in a browser? How nice would it be if we can write websites in Python? PyScript makes it possible! PyScript allows users to write Python in the browser. In this talk, I will introduce PyScript and discuss what does PyScript mean for data scientists, how PyScript might change the way data scientists work, and how PyScript can be incorporated into the data science workflow.

Can you suggest 2-3 topics for post-discussion?
Visualization, Model Deployment

3: 30 PM
-
4:00 PM EST

Sensitivity and Interpretability of AI-Models
Hanieh Arjmand, ML Researcher, Lydia.ai & Spark Tseung, Applied Data Scientist, Lydia.ai

About the Speakers:
Hanieh Arjmand is a Machine Learning Researcher at Lydia.ai where she focuses on discovering and applying the best machine learning techniques to healthcare and insurance problems to help insurers use machine learning to protect more people.
Spark Tseung is an Applied Data Scientist at Knowtions Research where he focuses on building frameworks for actuarial and underwriting validation to help insurers use machine learning to protect more people. Spark is working towards his Ph.D. in Statistics and specializes in the application of machine learning methods in Property & Casualty loss modelling and risk selection.

Which talk track does this best fit into?
Case Study

Technical level of your talk?
(Technical level: 4/7)

What you’ll learn:
Using case studies from our work, we will discuss potential approaches for designing some of the sensitivity tests, which have helped us understand different aspects of model behaviours and data biases.

Abstract of Talk:
Model interpretability is important especially in regulated industries where risk-sensitive decisions typically require transparency and reliability of the underlying model. While often model interpretability gets sacrificed in other fields in order to achieve superior predictive performance, this is not the case in the regulated industries such as healthcare where model fairness plays an important role. In this talk, we will present case studies to illustrate the importance of sensitivity analysis for model interpretability and to showcase our design and implementations. Depending on the use cases of machine learning models, sensitivity tests have to be specifically and carefully designed and implemented. Using our machine learning models on electronic health record (EHR) and human activity, we will discuss potential approaches for designing some of the sensitivity tests, which have helped us understand different aspects of model behaviour and even uncover the unwanted biases and behaviours that had to be eliminated.

NOVEMBER 30th SESSIONS

9: 45 AM -	Recommender Systems at Loblaw Digital <i>Kai Luo, Senior Applied Scientist, Loblaw Digital</i>
9:50 AM EST	About the Speaker: Kai works as a Senior Applied Scientist at Loblaw Digital, leading development of the core recommender systems used in personalization use cases across all lines of business. Prior to that, he completed a master’s degree at the University of Toronto, with a thesis relating to conversational recommendation. Abstract of Talk: Loblaw Companies Ltd is a retailing conglomerate that is the largest grocery retailer in Canada. Its subsidiaries include over 20 supermarket banners, such as Real Canadian Superstore, No Frills, and T&T. E-commerce, particularly for grocery, has become a significant part of the business , and personalization use cases play an important role in that domain. In this talk, Kai will discuss challenges relating to modeling customers’ behaviors and grocery products’ latent representation, and how we iterate our system to solve these challenges.
9: 52 AM -	The transformative role of AI/ML in heavy industries <i>Rex Lam, Director, Machine Learning Platform, Autodesk</i>
9:57 AM EST	About the Speaker: Rex Lam leads the Machine Learning Platform team to build platform capabilities that enable full ML cycle development and operational tools at Autodesk that aim to enable ML solutions faster, trusted and scalable. Abstract of Talk: AI/ML represents opportunities for Autodesk to drive insights & innovative solutions in architecture, design, and manufacturing tools.
9: 59 AM -	Are AI Scientists the key to organization's AI success? <i>Dr. Jodie Lombana, Founder and CEO, Tera Tera Collaborations Inc.</i>
10:04 AM EST	About the Speaker: Dr. Jodie Lobana is an experienced director, educator, and a researcher, with an integrated expertise in governance of artificial intelligence, risk management, internal audit, and human resources. Dr. Lobana is currently writing a book titled “Holistic Governance of Artificial Intelligence”, based on her Ph.D. Dissertation (to be released in Jan 2023). She is the founder and CEO of a startup named TeraTera Collaborations Inc., with a focus on training and consulting in the area of governance of artificial intelligence. Dr. Lobana has taught at various universities including McMaster University, York University, University of Waterloo, and Indian Institute of Technology. She is the Chair of the Advisory Board of the McMaster Artificial Intelligence Society, and a member of International Internal Audit Standards Board, as well as OECM Board. Dr. Lobana has been awarded "2019 Mary Keyes Award for Outstanding Leadership and Service to McMaster", and has been named "McMaster Women in Tech Aug 2020 Changemaker". Abstract of Talk: When companies think about how to enhance their AI success, they think about getting the best AI Scientists or Data Scientists. They think about data that they are going to need to develop AI. Although technical people and data are important for AI success, some other things matter more. These things are found on the upper floors of the organization, and if used properly, can significantly speed up an organization’s AI success. What are those things? The answer to this can be found in this lightning talk. Come join us and learn more.
10:13 AM -	AI & Sustainability: A \$50 trillion opportunity <i>Andrea Ruotolo, Global Head, Sustainability / ESG, Rockwell Automation</i>
10:18 AM EST	About the Speaker: Andrea is the Global Head of Sustainability/ESG at Rockwell Automation, the world’s largest industrial automation company, with responsibility for advancing innovation in sustainability for Rockwell’s customers, which include Fortune 100 companies in energy and manufacturing, representing millions of employees and hundreds of billions of dollars in annual revenues. Andrea is a passionate evangelist for the role AI/ML can play in dramatically improving the sustainability of the industrial sector. Across her nearly two decades of experience in leading technology innovation, in a career spanning Europe, Asia, and the Americas, Andrea has held multiple senior executive roles focused on applying advanced technologies to solve the sustainability challenge. She has served as co-founder and entrepreneur in smart grid consulting, global lead in the world’s largest engineering services firm in the energy sector, and senior director at a major utility. As well as her Fulbright Doctorate in an ESG analysis of sustainable energy systems, Andrea holds a B.A. from the University of la Plata in Argentina, a M.Sci. in Aeronautical and Aerospace Engineering from Madrid Polytechnic, and certification in Digital Business Strategy and AI from MIT Sloan School of Management. Abstract of Talk: The financial and business community have already caught on to the essential importance of sustainability. Investors now call for better practices and reporting on Environment, Social, and Governance, or ESG performance metrics. According to Bloomberg Intelligence, growth in ESG investing is fast becoming the new norm, with ESG investments projected to exceed USD 50 trillion by 2025 – more than 1/3 of all global assets under management. This movement of funds to ESG represents a massive, once-in-a-generation transition to an entirely new economy. Sustainability is incredibly complex, involving billions of moving parts and decisions. It starts at the edge, where exabytes of data are flowing from real-time sensors and controls in factories and power plants, which aggregate up to the top-level decision makers in companies, which aggregate up to the massive funds that hold portfolios of those companies, and to government regulators and policymakers. AI is critical in analyzing those exabytes of data and enables closed-loop optimization to reduce energy, water, and waste. In this session, we'll explore the top 3 needs and opportunities for AI to catalyze change toward more sustainable companies, economies, and societies.

10:55 AM

11:30 AM EST

Artificial Intelligence And Digital Pathology: Making The Most of Limited Annotated Data
Anne Martel, Professor, **University of Toronto**

About the Speaker:
Anne Martel is a Professor in Medical Biophysics at the University of Toronto, the Tory Family Chair in Oncology at Sunnybrook Research Institute, and a Faculty Affiliate at the Vector Institute, Toronto. Her research program is focused on medical image and digital pathology analysis, particularly on the development of self-supervised and weakly supervised methods for segmentation, diagnosis, and prediction/prognosis. In 2006 she co-founded Pathcore, a software company developing complete workflow solutions for digital pathology. Dr Martel is an active member of the medical image analysis community and is a fellow of the MICCAI Society which represents engineers and computer scientists working in this field. She has served as board member of MICCAI and is currently on the editorial board of Medical Image Analysis, on of the leading journals in the field.

Which talk track does this best fit into?
Advanced Technical / Research

Technical level of your talk?
(Technical Level: 6/7)

What you'll learn:
Self-supervision and smart sampling strategies are essential in digital pathology

Abstract of Talk:
Obtaining large datasets with detailed annotations for medical imaging AI projects is a time consuming and expensive process as it usually requires the input of expert radiologists and pathologists. Collecting data to train outcome prediction models is even more challenging as the number of patients with both imaging and follow up data may be small, and only weak labels are available. This talk will describe several semi-supervised and self-supervised approaches which can make more efficient use of small and/or weakly labelled datasets. The focus will be on digital pathology but the methods described are applicable any medical imaging modality.

10:55 AM

11:30 AM EST

Transforming The Retail Industry with Transformers
Rohit Saha, Applied Research Scientist & Kyryl Truskovskyi, ML Engineer, **Georgian**

About the Speakers:
Kyryl has over eight years of experience in the field of Machine Learning. For the bulk of this career, he has helped build machine learning startups, from inception to a product. He has also developed expertise in choosing and implementing state-of-the-art deep learning architectures and large-scale solutions based on them.

Rohit Saha is currently an applied research scientist at Georgian's R&D team and is assisting portfolio companies with their research endeavours. Owing to previous roles, he has experience building end-to-end machine learning pipelines. He holds a master's degree from the University of Toronto, and his research interests include generative modelling and transfer learning for Computer Vision tasks

Which talk track does this best fit into?
Case Study

Technical level of your talk?
(Technical level: 6/7)

What you'll learn:
The insights and findings we share in this talk are derived from using the latest ML techniques and tools for solving real-world use cases at SPINS and are not readily available on the internet. We also open the stage for Q&A at the end of this talk to address the questions from the audience.

Abstract of Talk:
In recent years, we have seen amazing results in artificial intelligence and machine learning owing to the emergence of models such as transformers and pretrained language models. Despite the astounding results published in academic papers, there remains a lot of ambiguity and challenges when it comes to deploying these models in the industry because 1) troubleshooting, training, and maintaining these models is very time and cost consuming due to their inherent large sizes and complexities 2) there is not yet enough clarity about when the advantages and challenges of these models outweigh classical ML models. These challenges are even more severe for small and mid-sized companies that do not have access to huge compute resources and infrastructure. In this talk, we discuss these challenges and share our findings and recommendations from working on real-world examples at SPINS, a data/tech company focused on the natural grocery industry. More specifically, we describe how we leverage state-of-the-art language models to seamlessly automate parts of SPINS' data ingestion workflow and drive substantial business outcomes. We provide a walk-through of our end-to-end MLOps system and discuss how using the right tools and methods has helped to mitigate some of these challenges. We also share our findings from our experimentation and provide insights on when one should use these massive transformer models instead of classical ML models. Considering that we have a variety of challenges in our use cases from an ill-defined label space to a huge number of classes (~86,000) and massive data imbalance, we believe our findings and recommendations can be applied to most real-world settings. We hope that the learnings from this talk can help you to solve your own problems more effectively and efficiently!

11:35 AM

12:05 PM EST

Creative Outputs and Machine Learning
Jordan Shaw, Creative Technology Lead, **Half Felix**

About the Speakers:
Jordan Shaw is an artist, and creative technologist raised and is currently based in Toronto, Canada. He grew up in Scarborough and received his MFA from OCAD University's Digital Futures program leading to his thesis being exhibited during Vector Festival at InterAccess. Before that, he completed his undergraduate degrees at Carleton University and Algonquin College, where his final installation was exhibited and recognized during ACM SIGGRAPH.

His work is related to exposing the hidden and unseen aspects of technology and the digital environment around us. The manifestation of this work tries to visualize the hidden interactions between people and technology, data collection and these digital systems trying to understand the physical world.

Jordan has exhibited internationally in Australia, Canada, Germany, Spain and the United States of America.

Which talk track does this best fit into?
Technical / Research

Technical level of your talk?
(Technical level: 4/7)

Are there any industries (in particular) that are relevant for this talk?
Computer Software, Marketing & Advertising, Art + Technology // Cultural Reflection and

What are the main core message (learning) you want attendees to take away from this talk?
The exploration, evolution and progression of ML throughout the last couple of years through a creative lens.

Abstract of the talk:
A creative evolution of ML in the arts. The speaker started producing artwork and exhibiting ML pieces in 2015. The reflective perspective of progression in ML and AI while observing creative model biases provides a unique perspective for the future of ML in creative fields and may influence popular culture down the road.

11:35 AM

12:05 AM EST

Scaling & Evolving the Machine Learning Platform at DoorDash.
Hien Luu, Head of Machine Learning Platform, DoorDash

About the Speaker:

Hien Luu is a Sr. Engineering Manager at DoorDash, leading the Machine Learning Platform team. He is particularly passionate about the intersection between Big Data and Artificial Intelligence. He is the author of the Beginning Apache Spark 3 book. He has given presentations at various conferences such as GHC 2022, Data+AI Summit, XAI 21 Summit, MLOps World, YOW Data!, appy(), QCon (SF,NY, London).

Which talk track does this best fit into?

Case Study

Technical level of your talk?

(Technical level: 5/7)

Who is this presentation for?

Senior Business Executives, Product Managers, Data Scientists/ ML Engineers and High-level Researchers, Data Scientists/ ML Engineers, ML Engineers

What you'll learn:

The journey in scaling the ML platform

On a scale of 1-10 how mature is this applied AI application you plan to discuss?

10/10

Pre-requisite Knowledge:

High level understanding of microservices

What kind of DevOps tools you plan to discuss? Open source?

CI/CD, Git, MLFlow,

What are some of the languages you plan to discuss?

Python, Kotlin

What are some of the infrastructures you plan to discuss?

Feature engineering at scale, low latency and high QPS model prediction service

What is unique about this speech, from other speeches given on the topic?

This is a case study about our journey of building ML platform at DoorDash

Abstract of Talk:

As DoorDash business grows, the online ML prediction volume grows exponentially to support the various Machine Learning use cases, such as the ETA predictions, the Dasher assignments, the personalized restaurants and menu items recommendations, and the ranking of the large volume of search queries. In this session, we will share our journey of building and scaling our Machine Learning platform and particularly the prediction service, the various optimizations experimented, lessons learned, technical decisions and tradeoffs made. We will also share how we measure success and how we set goals for the future.

Can you suggest 2-3 topics for post-discussion?

Adopting MLOps

11:35 AM

12:05 PM EST

Interpretability Tools are Feedback Loops
Jinen Setpal, Machine Learning Engineer, DagsHub

Which talk track does this best fit into?

Technical / Research

Technical level of your talk?

(Technical level: 6/7)

Are there any industries (in particular) that are relevant for this talk?

Computer Software, Researchers within Academia

Who is this presentation for?

Data Scientists/ ML Engineers, ML Engineers, Researchers

What you'll learn:

Most guides on Interpretable Machine Learning, while focusing on the seemingly black-box nature of models, fail to address the challenge of it only working post-training. It's addressed as a final check tool to ensure it isn't learning tragically differently from what the researcher expects, but not directly as part of the training feedback loop. I aim to address this.

What are the main core message (learning) you want attendees to take away from this talk?

If, by the end my presentation, attendees are able to identify techniques for applying the proposed approach within their internal systems, or find themselves motivated to further research the ideas presented, I'd consider the talk a success.

Pre-requisite Knowledge:

It's a technical presentation, and requires participants to be familiar on a high level with model optimization techniques and machine interpretability. They should also be very familiar with the standard classification pipeline of a feedforward neural network.

What is unique about this speech, from other speeches given on the topic?

I work extensively on ML Reproducibility. In fact, it is the fundamental work done within my research grant funded by Google. Time and again, I find that papers tend to document the end-result and the problem statement extensively, but not everything in between. That's where the real learning takes place, understanding what DIDN'T work and WHY.

This is rarely ever documented. Papers list working parameters, both functions and hyperparameters, but fail to explain why. I've struggled a lot with this and hope to relay potential solutions through the extent of my presentation.

Abstract of Talk:
Fundamentally – Machine Learning as a field is designed to emulate the way humans think; hence, *neural* networks. When we train our models, we use optimizers and loss functions to measure their success. While these functions make sense mathematically, they are far from intuitive or explaining what happened behind the scenes. It’s hard to pick the correct functions, and performing huge grid searches to hyperparameter tune at scale is as logical as bruteforcing an SHA-256 hash.

On the other hand – Interpretability techniques can’t really be used in a training context but are intuitive in helping us understand how a given model interprets a set of data.
This talk aims to bridge the gap between the two, connecting them within a single training loop to maximize training effectiveness without disproportionately increasing compute or training time. Making training intuitive to how humans learn should help develop models that actually work, without resorting to “useless” training.
I aim to showcase – with a practical demonstration – learning techniques to build feedback loops wherein interpretability is used to better optimize a training sequence. I also aim to discuss how this carries forward to complex architectures, and a potential approach for their relevant implementation.

Structurally: the talk would provide an overview on machine interpretability, provide a brief overview on optimizers and loss functions before jumping into the implementation walkthrough of a case study. The case study uses TensorFlow, but can generally be applied to any desired framework.

Can you suggest 2-3 topics for post-discussion?
Information Retrieval, Explainable AI, relevance of the above in product development

1:15 PM
-
1:45 PM EST

2022 AI Index Report Briefing
*Nestor Maslej, Research Manager, **Stanford Institute for Human-Centered Artificial Intelligence (HAI)***

About the speaker:
Nestor Maslej is a Research Manager at Stanford’s Institute for Human-Centered Artificial Intelligence (HAI). In this position, he manages the AI Index and Global AI Vibrancy Tool. Nestor also leads research projects that study AI in the context of technical advancement, ethical concerns and policymaking. In developing tools that track the advancement of AI, Nestor hopes to make the AI space more accessible to policymakers.

Nestor also speaks frequently about trends in AI. He has delivered presentations about the AI Index to teams at the World Economic Forum, Centre for Data Ethics and Innovation and Global Arena Research Institute. Nestor has also testified to the Canadian Parliament’s House of Commons Standing Committee on Access to Information, Privacy and Ethics on the use and impact of facial recognition technology in Canada.

Prior to joining HAI, Nestor worked in Toronto as an analyst in several startups. He graduated from the University of Oxford in 2021 with an MPhil in Comparative Government, where he used machine learning methodologies to study the Canadian Indian Residential schooling system and Harvard College in 2017 with an A.B. in Social Studies.

Technical Level: (3/7)

Which talk track does this best fit into?
Technical / Research

What are the main core message?
That AI is here in a way that it was not before, and that as a society, we need to think critically about the role AI should play in our lives.

Are there any industries (in particular) that are relevant for this talk?
Wide ranging: given that AI will be ubiquitously deployed I figure all industries can benefit from the presentation to a degree.

What you'll learn:
That AI is here in a way that it was not before, and that as a society, we need to think critically about the role AI should play in our lives.

Abstract:
Learn about some of the main trends in AI, as told to you by the 2022 AI Index Report. The AI Index is one of the most widely read annual reports on trends in AI and has informed AI policymakers and industry leaders across the globe. This presentation covers some of the main trends explored in the report, namely trends in areas such as research and development, technical advancement, ethics, economics, policy and education.

1:15 PM
-
1:45 PM EST

Marketplace Experimentation at SpotHero
Amish Popli, Data Scientist, SpotHero

About the Speaker:
Amish Popli is passionate about solving challenging business problems using data science and machine learning. He supports multiple departments at SpotHero including, but not limited to, marketing, sales, and product development. He likes data, manipulating it, making it (simulation), modelling it, visualizing it, and yes, even cleaning it. He works with different PMs and engineers in different domains and has brought many successful products from discovery to production.

Technical level of your talk?
(Technical level: 4/7)

Are there any industries (in particular) that are relevant for this talk?
Parking

Who is this presentation for?
Product Managers, Data Scientists/ ML Engineers

What you’ll learn:
Experimentation is a very nebulous topic. There is a lot of companies, articles and research available on google, but each company has its own unique way of running and measuring impact from experiments.

Pre-requisite Knowledge:
Knowledge of basic statistical tests

What is unique about this speech, from other speeches given on the topic?
In my knowledge there is no company in North America doing data science/ML in parking industry. The problems we are solving are present in other industries but parking adds another layer of complexity on top of it.

Abstract of Talk:
SpotHero is the biggest and fastest growing off-street reservation platform in North America. It is a two sided marketplace involving drivers and parking garage owners. The data science team at SpotHero is working on many interesting problems in the areas of dynamic pricing, marketing, ranking etc. One of the key challenges that we face is how we test our machine learning models in production and make sure that the changes we make lead to an improvement in our KPI’s. In this talk, I will focus on how SpotHero runs experiments whenever we make improvements or create a new model to generate prices for our parking spots. I will cover why the general A/B test framework will not work in our scenario, various approaches that we considered and introduce switchback experimentation as an alternative. I will discuss our experiment design and conclude the talk with a result from one of our experiments and our technical architecture.

Can you suggest 2-3 topics for post-discussion?
A/B Tests, Switchback experiments, challenges in running live experiments

1:50 PM
-
2:20 PM EST

Assessing Alignment of Climate Disclosures Using NLP for the Financial Markets
*Quoc Tien Au, Data Scientist and Aysha Cotterill, Data Analyst, **Manifest Climate***

About the Speaker:
I am a data scientist at Manifest Climate, working on applying machine learning and natural language processing to climate disclosures. Extracting information at scale is paramount to increase transparency in financial markets, so that we can improve decision-making with data-driven climate information.

Which talk track does this best fit into?
Case Study

Technical level of your talk?
(Technical level: 5/7)

What you'll learn:
How an early-stage startup runs machine learning experiments ; takes decisions balancing model performance, model explainability, resource constraints and added business value ; uses deep language models to create the most valuable business opportunities.

Abstract of Talk:
Climate-related disclosure is increasing in importance as companies and stakeholders alike aim to reduce their environmental impact and exposure to climate-induced risk. Companies primarily disclose this information in annual or other lengthy documents where climate information is not the sole focus. To assess the quality of a company's climate-related disclosure, these documents, often hundreds of pages long, must be reviewed manually by climate experts. We propose a more efficient approach to assessing climate-related financial information. We construct a model leveraging TF-IDF, sentence transformers and multi-label k nearest neighbors (kNN). The developed model is capable of assessing alignment of climate disclosures at scale, with a level of granularity and transparency that will support decision-making in the financial markets with relevant climate information.

1:50 AM
-
2:20 PM EST

Latent User Intent Modeling in Recommender Systems
Bo Chang, Software Engineer, Google Brain

About the Speaker:
Bo Chang is a software engineer at Google Brain, based in Toronto, Canada. Prior to that, he was a machine learning researcher at Borealis AI. He finished his Ph.D. in statistics at the University of British Columbia.

Which talk track does this best fit into?
Advanced Technica I/ Research

Technical level of your talk?
(Technical Level: 7/7)

What you'll learn:
How to better model user intent in recommender systems using a latent variable model.

Abstract of Talk:
The current sequential recommender systems mainly rely on users' item-level interaction history to capture topical interests and lacks a high-level understanding of user intent. It is challenging to explicitly define and enumerate all possible user intents. We propose to use latent variable models to capture user intents as latent variables through encoding and decoding user behavior signals, with an application to a large industrial recommender system.

1:50 PM
-
2:20 PM EST

Managing AI and the Associated Risks in a Regulatory Environment
David Van Bruwaene, Founder and CEO, Fairly AI & Fion Lee-Madan, Co-Founder and COO, Fairly AI & Susie Lindsay, Counsel, Law Commission of Ontario

About the Speakers:
Fion has over 20+ years of experience in enterprise software as a Solutions Architect (ex-Sapient, ex-Intuit, ex-ATG – acquired by Oracle.) She double majored in Computer Science and Human Biology at the University of Toronto and has an MBA from Boston University. She is a technical committee member of the CIO Strategy Council of Canada. She is a champion of DE&I, and a major supporter for women in tech as both a mentor and coach. She guest lectures for AI Ethics at Lighthouse Labs, an education company with a goal to bring more diversity into the Data Science field.

David has developed a deep understanding of ethics and formal logic, model theory, and Natural Language Processing (NLP) throughout his career in business and academia. He taught Business Ethics at the University of Waterloo and graduate level Ethics at Cornell and is a sought-after speaker for AI Ethics, Compliance and Risk Management at conferences around the world. On an exchange scholarship from Cornell to Berkeley, he became fascinated with powerful Natural Language Processing. David applied this background to cyberbullying and related NLP in his first AI startup ViSR (acquired by SafetoNet) where he was the Head of Data Science, then became the CEO and Board Member. His unique perspective as a practicing Data Scientist to being the CEO and Board Member at the top made David conscious of tensions between technical vs business decisions and the impact on people’s lives in the resulting automation of decision-making.

Susie is Counsel at the Law Commission of Ontario where she leads numerous LCO projects including: AI and the Civil Justice System, Protection Orders, and the LCO’s joint initiative on AI and Human Rights with the Ontario Human Rights Commission and Canadian Human Rights Commission. Before joining the Commission Susie practised regulatory law at a large communications company, and civil defence litigation at a boutique litigation firm. Susie is a graduate of Queen’s Law School, a Fulbright Scholar, has a Master of Laws from Harvard Law School and was a fellow at the Berkman Klein Centre for Internet & Society at Harvard University.

Which talk track does this best fit into?
Business Strategy

Technical level of your talk?
(Technical level: 2/7)

What you’ll learn:
Latest expert views on managing AI risks in the fast changing regulatory environment

2:25 AM
-
2:55 PM EST

Solving Product Substitutions, A Big Problem in Grocery E-Commerce – Through Self-Supervised ML
Jawad Ahmed, Staff Applied Scientist, Loblaw Digital

About the Speaker:
Jawad currently works as a Staff Applied Scientist at Loblaw Digital, supporting ML teams building personalization and recommender systems for different lines of business of Loblaw Companies. He has 8 years of industry experience in Applied AI/ML. Previously, he worked at Flipp, Dialpad and McKinsey Solutions. His areas of interest are using ML research applications to help build products with scalable ML solutions in NLP, Conversational AI, Computer Vision and Recommender Systems.

Which talk track does this best fit into?
Case Study

Technical level of your talk?
(Technical level: 5/7)

What you’ll learn:
The talk covers the data curation process by which we prepared a benchmarking Products Substitutions dataset using historical human-selected substitutions data at Loblaw. The audience will learn about self-supervised ML approaches we used to recommend product substitutions benchmarked against the above mentioned products substitutions Testset.

Abstract of Talk:
Background: Loblaw Companies Ltd is the largest grocery retailer in Canada. It operates multiple popular banners with Real Canadian Superstore, No Frills, and T&T being some of the most popular ones. E-commerce of grocery has become a significant part of the business accounting for more than \$2 billion in sales per year. Problem: Shopping for groceries online is an inherently different process than shopping in person. We take for granted the in-store shopper’s ability to make quick decisions on the fly when faced with the issue of product availability. We fulfill from stores to ensure freshness which has a very dynamic inventory. This makes promises of items collected, sometimes a day or two after the order depending on the customer’s delivery date, affected by many factors – some of which we cannot control. Thus, we need a solution to substitute items that are out of stock at the time of picking to make sure the customer experience is minimally impacted. While shopping at a physical store, a customer can make a suitable choice of an alternative. In the e-commerce process of grocery shopping, either the customer has to make a selection of the substitute, or the Loblaw employee picking the order on behalf of the customer needs a relevant suggestion on the best substitute for the given item, personalized for the given customer. Loblaw has historical data available on what selection was made by customers from the list of various possible substitute options available for a given item. Additionally, there is data available on the choices made by pickers – the employees who shop at the store to fulfill customers’ orders. This provides us an opportunity to tailor product similarities toward product substitutions that are tied to business metrics. Solution: We explored multiple solutions to solve this problem. Our most promising solution that we wish to present leverages features extracted from text descriptions and images of products. In this talk, we will discuss how our approach evolved over time and how this cutting-edge self-supervised method is a big improvement over the traditional techniques.

2:25 AM
-
2:55 PM EST

Data FAILS
Nicolás Venegas Oliva, Technical Lead of Advanced Analytics, LATAM Airlines
Sarah Sun, Director Data Science, Scotiabank

About the Speaker:
Nicolas Matias Venegas Oliva has 2 years of experience in backend development, 2+ years in data processing and the last 3+ years as Advanced Analytics technical leader at LATAM Airlines. During this time the team has grown from 9 to 48 highly trained professionals. It has also become the team with the highest impact generation within the company and a reference in the region in terms of MLOps and measured business impact through data products.

Sarah Sun - A decade in data has taken Sarah across multiple industries, including banking, technology, and natural resources. While specializing in data strategy, she was trained as a data scientist and has worked across the industry in innovation, governance, AI, and also a stint as CEO of a startup. Working in data has taught Sarah some value lessons - everything from seizing opportunities, the important of mental health, and the power of sharing stories. Sarah was named one of the Women Executive Network's Top 100 Most Powerful Women in 2019

Which talk track does this best fit into?
Business Strategy

Technical level of your talk?
(Technical level: 1/7)

Are there any industries (in particular) that are relevant for this talk?
Applicable for any industry!

What are the main core message (learning) you want attendees to take away from this talk?
Lessons to be learnt in the experiences shared on what not to do. In addition, we like to talk about the wins, but the reality is fails are more frequent, but also more lessons to be learnt!

Abstract of Talk:
We like to talk about the successes...but why don't we ever talk about the FAILS across the data world? Join us as we swap stories crossing data, industrial, and even geographical boundaries. We may have failed...but maybe amongst all the tales there's a lesson or two to be learnt across networking, recruiting, planning, model building, engineering....you name it ;)