Individual Assignment : York University CSML1000 : Machine Learning in Business Context

Gouri Kulkarni, student # 218491811

07/11/2020

1. Unemployment. What happens after the end of jobs?

Prehistorically, division of labour arose from differences in age and sex. In family heirarchies and tribal heirarchies, the young hunted for food, the older were educators, the women gathered food and reared children. With the development of pottery, textiles, agriculture and metallurgy was born the need for specialization and heirarchies became more complex. This led to distinct social and economic classes like priests, merchants and workers. In the pre-industrial age, certain industries grew at a faster pace than the others. Systems were developed. Factories were formed. *Adam Smith (1776)* described the new production system in "Wealth of Nations". The invention of the assembly line by *Henry Ford (1914)* was a major milestone the development of the heirarchy of labour. Jobs that were traditional then were replaced with specialized jobs requiring skills working on the assembly line.

Though this is not a write up on Evolution, it is worth while to mention that the development of the organization of work differentiates humans from animals and is responsible for starting the human conquest of nature. Humans have a complex brain structure, are blessed with linguistic communication and have been developing tools since *homo sapiens*. From this was born the *economy of labour*.

Today, *ML and AI* are emerging technologies as much as the assembly line was once upon a time. The invention itself, though originating in America spread all over the world and helped countries become self sufficient in manufacturing and exporting.

Increased productivity led to the use of women and children as inexpensive labour. It is argued that division of labour destroyed skill. Automation made some work dull and repetitive. This led to new social and political movements. With colonization came worldwide division of labour in the form of slavery.

How do we ensure the advances in *ML and AI* do not produce the same side effects? Back then, the effects were gradual and in line with the pace of life. The effects varied across the world.

Today, we live in the data age.The heirarchy of labour is more complex than ever. If self driving trucks become available in the next decade, we will not be surprised to see alternate sources of employment arise in parallel jsut as the assembly line saw the creation of a new set of jobs and skillsets. *COVID-19* has escalated the pace at which technology is evolving and is shaping the economy of labour. We are more than ready to embrace self serve options. *AI and ML* are the next major technology after the assembly line and the internet to revolutionize the economy of labour. While it is believed that *AI and ML* will put millions out of work , it will be a gradual transition. We are already seeing the gig economy become more common place than full time jobs.Entrepreneurship is on the rise.

Will this emerging technology replace all jobs? The nature of work will change. Take for example restaurants. We are already seeing robots do the work of waiting tables.

Looking back, we can ridicule over the whole idea of spending our time to live. However the brains behind the development of the robots , driver less vehicles and drones are also people with a different skillset - data analysis , engineering, research. *AI and ML* technologies can be used to effectively re-skill employees. It is said that wealth inequality between countries will go up. AI Governance is a mammoth task, but if we handle

Governance well, we can ensure balanced distribution of wealth. Data Governance at a macro level is now AI Governance.

A person's occupation has been their identity for centuries. With AI and ML technologies freeing up our time , we will see ourselves leaving the repetitive, monotonous, predictable tasks to machines. People will have time at their disposal to perform work at a higher level. Over all, the quality of jobs will improve. A garbage collector operating a robotic arm and vehicle will become commonplace.

We are still a far cry from promising every human the benefits of this emerging technology. A universal basic income is one of them.

In Isaac Asimov's words (1988) : "The saddest aspect of life right now, is that science gathers knowledge faster than society gathers wisdom." How true. We are drowning in data, but starved for wisdom.

The growth of AI research and applications has outpaced the development of regulatory frameworks and standards around the world.

Humans can endeavor to be better than robots. Humans will team up with machines.

Industry 4.0 is the revolution we are experiencing right now and we are just starting to see the magnitude,scope and challenges.

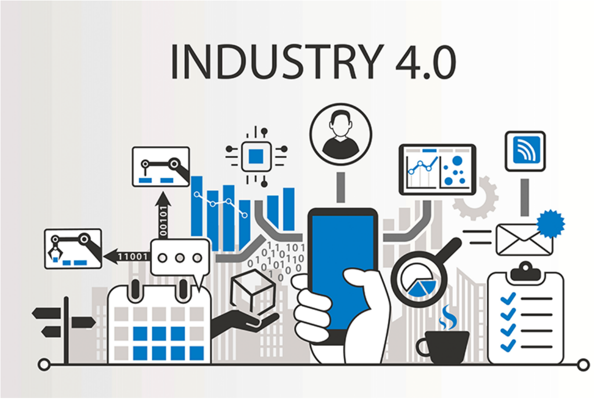
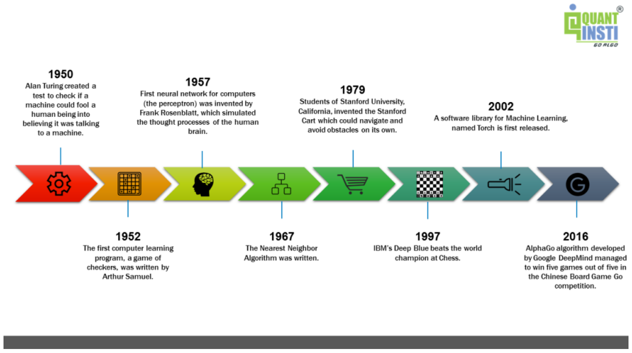


Figure 1: Industry 4.0

A brief timeline of machine learning



1950
Alan Turing created a test to check if a machine could fool a human being into believing it was talking to a machine.

1952
The first computer learning program, a game of checkers, was written by Arthur Samuel.

1957
First neural network for computers (the perceptron) was invented by Frank Rosenblatt, which simulated the thought processes of the human brain.

1967
The Nearest Neighbor Algorithm was written.

1979
Students of Stanford University, California, invented the Stanford Cart which could navigate and avoid obstacles on its own.

1997
IBM's Deep Blue beats the world champion at Chess.

2002
A software library for Machine Learning, named Torch is first released.

2016
AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Chinese Board Game Go competition.

Figure 2: Machine Learning timeline

Machine learning has evolved over the years. Today the rate at which solutions are being developed is growing exponentially with the availability of computing power and infrastructure. This growth is outpacing the rate at which we are developing regulatory frameworks. A unified approach is required.

2. How do we distribute the wealth created by machines?

As discussed, in today's *labour economy* we sell our time for money. The speed at which work gets done has been getting better day by day. Will we sell less time for more money? Will having to spend less time at work mean that we get quality time with family? Or will there be a differentiation between work and life? Ethical and responsible implementation of AI will improve the overall quality of life.

When we talk about inequality and the distribution of wealth created by machines, we are referring to the industry giants Google, Facebook, Microsoft, Amazon and Apple who are at the frontline of *AI and ML.*

Quotes from the mission, vision and values statements of each :

Google : Committed to significantly improving the lives of as many people as possible.

Facebook: Give people the power to build community and bring the world closer together.

Microsoft: Empower every person and every organization on the planet to achieve more.

Apple: To make a contribution to the world by making tools for the mind that advance humankind.

Amazon: Customer obsession rather than competitor focus, passion for invention, commitment to operational excellence, and long-term thinking.

Their adherence and perseverance to their statements and values has made them winners and leaders of the new AI driven economy. However, every individual is continuously feeding data in to the network of their warehouses. The data collectors stand to gain from the process of collection.

A hundred years back, with the invention of the automobile, thousands of jobs were created. The high cost of machinery used in factories could only be justified if owners of the machinery had employees using them for production. Here, the machinery had a demand. Employees hired by the owner of the machine were the users. The owners got richer while many earned their livelihood.

Today with AI and ML technology, employers need to hire less bodies and end up taking home more revenue.

It is unfair that the wealth created is not distributed equally. But how do we ensure equal distribution of future wealth ?

There are several propositions. A universal income. A fair post-labour economy would use data to generate statistics and fairly pay employees. The concept behind usage based models can be used to evaluate work and pay fairly.

Quoted from the World Economic Forum: *"Entrepreneurship is a fundamental driver of economic value and job creation, whether in the form of a new startup or as a regenerating force within an established company. Entrepreneurship is about acting under conditions of uncertainty, and in the absence of calculable risk. As they discover or exploit opportunities, and mobilize the resources necessary to act on them, Entrepreneurs unleash the forces of creative destruction in ways that transform existing industries and create entirely new ones."*

Entrepreneurship is on the rise around the world as we try to find better solutions to mankind's problems.

3. Humanity. How do machines affect our behavior and interaction?

Turing test: The Turing Test is a method of inquiry in artificial intelligence for determining whether or not a computer is capable of thinking like a human being. The test is named after Alan Turing, the founder of the Turing Test and an English computer scientist, crypotanalyst, mathematician and theoretical biologist.

AI solutions using reinforcement learning can be used to change human behavior. The humans who took the challenge against Alpha Go learned more about the game Go than in their entire life.

AI could be humankind's greatest inventions if used the right way. How do we build safe AI systems that solve problems and advance scientific discovery, positively impact human behavior and interaction? They can help us understand human behavior better. They can help us understand what we are thinking and what should be the next correct and logical step. Understanding emotions are the next frontier. Understanding emotions such as fear, honesty, anger can be useful in assessing risk and gaining insight. Uses in medicine are unlimited, for example to cure addictions and nurture healthy habits.

4.Artificial stupidity. How can we guard against mistakes?

Technologism is the belief in the power of technology to shape or improve human society. But is it actually improving or gradually deteriorating society?

It is because of advances in technology that humans are in charge of the world today and world population could grow at the exponential rate it is.

Humans have been the smarter species. We invented technologies. The human brain was capable of coming up with inventions. But the use of technology is making our facilities weaker.

Books are one of the earliest form of technology. Man found a use for paper and scrolls to record memories. Cash and currency are also a form of memory - a record of our wealth. Today we don't count cash and change but leave it to the systems to produce and tally. We are also blessed with information overload.

True, there are the AI researchers and practitioners today who are sharper and exercise their faculties to invent technology for the benefit of humankind. But the vast majority are end users and are already becoming lazy and out of shape with their abilities due to reliance on technology. Think of how much we rely on a GPS. Humans developed cartography and conquered the world. Will ths technology and AI help us further in our quest to conquer the world?

We train AI systems to solve problems and do things better than humans. We make sure the robot is at an intelligence level above humans. Deepmind engineers created the AlphaGo and challenged humans to compete against it. With Reinforcement learning,the robot could learn to make good sequences of decisions. A robot does not know how its decisions can affect the world. If the robot competed with a bad player or someone who did not know the game , whould it have learned lesser?
In video games, with the help of RL, the robot learns to play the game better than people.

Mankind is becoming dumber by ignoring commonsense and following algorithms. AI systems have their limitations and have instead been coined as Artificial Stupidity.

Racist Robots. How do we eliminate AI bias?

How does any AI system develop bias ? AI systems are trained on data. If the training data is distributed with bias towards any category , then the results of the model will display the bias.

Tay was a machine learning project started by Microsoft.
It mimicked the personality of a 19 year old girl. She learned from the conversations she had with people. Unfortunately, Microsoft did not foresee the risks involved with placing the bot on Twitter. People started tricking the bot with racial slurs that the bot learned from, and outputted inappropriate content. This wans an unintended outcome. Microsoft scrapped the project. The designers had skipped the part of teaching the bot what not to say. The bug was fixed in a later release.

Following lessons learnt from Tay and other failed projects, Google is using deep learning to reduce Conversational AI toxicity by releasing toxicity scores and using synthetic datasets.

Facial recognition has been used by the police force to identify and rate criminals. The software claims to help police catch dangerous criminals and make communities safer. But most often, the systems make wrongful decisions. People have been arrested based on their colour, gender and other demographic factors. The intention was to free up police officer's time looking up suspects. Wrongful arrests were the unintentional outcomes. Facial recognition systems are fuelled by large datasets. They can only make a detection based on the data they are trained on. If there is not enough variety in the data , predictions will not be accurate enough and could be wrong. Live AI software can be used to detect faces and pinpoint suspects. Tech companies have been boasting of the capabilities of their software. The software scans faces , measuring key features. The data fed to the model mostly coming from a certain demography, the model would not be able to accurately provide results on a test set.Following a lot of incidents where innocent individuals were targeted by police, Amazon, Microsoft and IBM froze the supply of facial recognition systems to police force until regulations were developed and implemented. The challenge with facial recognition algorithms is that bias can be spotted only when a wrongful detention or test case occurs.

It is up to humans to influence how AI evolves. AI systems automate repetitive tasks and free up humans to pursue more creative tasks. We cannot risk implementing AI systems if the model takes biased decisions. Here is an infographic showing how bias can creep into data .

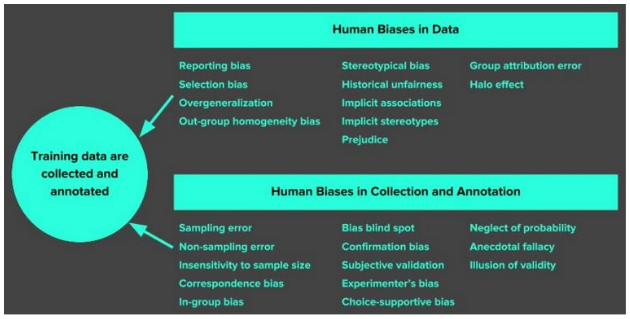


Figure 3: Human Biases in Data

Most often, what people share is not a reflection of real world frequencies, resulting in reporting bias. When our selections do not reflect a random sample, selection bias creeps in. When we see outgroup members as more alike than individual groups, it results in outgroup homogeneity bias.

It is important to have every possible group represented in the dataset. In reality, some groups might be represented less positively than others. Labels could be biased. There can be bias in model interpretation, and over generalization. Many times, correlation is linked to causation. We tend to rely on automation, rather than human decisions, leading to automation bias.

Bias in AI is a network effect, Human bias can creep in at any stage of a machine learning project : data collection, annotation ,model development, model training, model output. The output serves as further training data . Hence, bias gets amplified in the machine learning development process.

Predicting criminal behaviour Automation bias - overgeneralization, correlation fallacy algorithms predict ares to deploy officers . data models are trained on are where crimed have been reported. The algorithm predicts the future from past . other areas that might have had crime don't get explored.

Facial recognition:

There is an interesting study by CSIS (2020) that shows that majority of algorithms exhibit demographic differences in reporting of false positives. The effect was larger on false positives than false negatives. This means, a person could be wrongfully arrested if he/she belonged to an underrepresented demographic group. But as the accuracy of the algorithm improved, the effect reduced. It also showed that bias can be eliminated with the right processes and algorithms. The most important factor is the training data. The fact that facial recognition algorithms developed in the US performed poorly on Asian faces than facial recognition algorithms developed in China. Developing regulation proposals and conducting training data audits can help mitigate the risks.

We can avoid bias by understanding our data, combining data from multiple sources, using training and test sets from different distributions, having a hard set of test cases that the model should perform well on, and using hold out datasets.

When datasets are released, we must release more information on data collection , annotation so the public can understand the biases present in the dataset.

Some bias mitigation strategies are : removal of features to de-bias the data ,adding desired variables to increase model performance , use of proxy data ,use of adverserial multi task learning where we predict the main task and also the thing that we don't want to be predicted and then use negative gradient

Just as datasheets are used when releasing datasets, model cards must be used when releasing a model. The model card provides information on what it is, what it does, how it works, factors, developers, intended uses, metrics used to understand fairness, information about model evaluation, risk and benefits, limitations of evaluations and intersectional evaluation.

6. Security. How do we keep AI safe from adversaries?

Machine learning has penetrated every possible field- finance, healthcare, education, transportation are a few. A security breach of a machine learning application can have severe repercussions.

Just as cyber criminals, hackers and phishers make use of the computer or network as a tool or target to perform malicious activities such as spreading viruses, data theft, spamming or fraud there are individuals or organizations whose focus in not to used ML and AI for good but compromise the algorithms.

Militaries over the world are using armed drones to allow for remote delivery of lethal force. As much as a driverless vehicles can lower the risk of accidents, unmanned combat aerial vehicles lower the risk to soldiers. With no on board human pilot, they can be designed to have a lower weight and smaller size. But what about humanitarian concerns? The first armed drone strikes occurred in Afghanistan and Yemen (2001 and 2002). Since then , drones deployed to carry out attacks have resulted in civilian casualties. Models can go wrong and can never be 100% accurate. Computer vision plays a key role in detecting various types of objects while flying in mid air. High performance on board image processing and neural networks are used for object detection, classification and tracking.

What if the algorithms used by military drones were compromised ?

We have a framework to detect and respond to adversarial attacks against machine learning systems. The framework follows the style of ATT&CK,used by cyber security analysts. It was developed by MITRE. The adverserial ML threat matrix is a joint effort between AI researchers at 13 organizations including IBM, Microsoft, Nvidia and Mitre. * https://www.mitre.org/publications/project-stories/mitre-microsoft-others-take-on-machine-learning-threats*

Machine learning engineers should refer to the matrix to get a holistic view of security in emerging ML systems, causes and new risks. Each column of the matrix refers to one tactic such as evasion and each cell refers to a technique. How does on attack a machine learning system? With machine learning code being available open source, ad the APIs used to deploy the model, it is easy for attackers to gather information on the actual working of a model.

AI is an emerging technology. In parallel, there are emerging risks and threats. For example, a financial loan application recommendation system may be rule based today , but future versions could be neural network based.

Every new technology brought with it risks. Security tools have been developed to detect DDoS attacks and the like. The challenge with detecting threats to ML systems is that the malicious behaviour could be embedded in one of millions of parameters of neural networks.

Model developers and machine learning researchers should pay more attention to the security aspect of the model and look beyond a single performance metric such as accuracy.

7. Evil genies. How do we protect against unintended consequences?

"Why the sea is salty" is a classic Korean tale. The sea was not always salty. A thief steals a king's magic millstone which turns every wich true. The mill obediently provide what ever is demanded. But the thief forgets the command to stop the mill from procusing more. Once after ordering the mill to grind salt, the thief forgets the command to make it stop. That is why the sea is salty.

In Stuart Russel's words, "If we cannot switch a machine off because it won't let us, we're really in trouble."

When demanding our AI systems, we may not be careful enough what we wish for. The difference and bigger threat is that AI affects billions of people.

An AI system can only deliver the goal set out for it. Youtube's recommendation algorithms are an example. Not all recommendations are a match. Recommendations could be no where near anticipated and could actually be spreading misinformation. This is reality, different from the intent. The burden currently lies on ML and AI designers to decide what consequences a system could have. This is much like product development.

What experience does one want out of driverless vehicles? Colission avoidance is a top requirement. In the quest to avoid colissions, a driverless vehicle could brake excessively or drive too slow, unlike when a human drives.

Artificial neural networks using reinforcement learning were developed , which learnt from scratch how to play and beat Atari games. With reinforcement learning , an AI system learns to optimize it's reward function. On the other hand, inverse reinforcement learning tries to learn what reward function a human is optimizing.Given a set of actions, it deciphers the underlying goals. If AI systems can quantify their own uncertainity about the human's preferences, the robot can gauge when to safely act. This can be achieved with deep learning. But human preferences change, and humans can have different short term and long term goals. Our actions do not live up too our ideals. These are the challenges. Another challenge is, if the owner is bad , will the robot pick up the preferences of the owner ? This is where values come in the picture. Robots should be nurtured just like a good parent will bring up children.

8. Singularity. How do we stay in control of a complex intelligent system?

Another differentiating factor between humans and animals is intelligence. Humans possess the intelligence that animals do not.

IBM developed the Deep Blue in 1997. The Deep Blue was programmed to play chess against Garry Kasparov. At that time, we thought that AI had finally arrived and that this was the ultimate. Deep Blue was programmed only to play chess. There were clear rules defined. This was symbolic AI. To use the same idea to play different games would have been a meticulous task. As computing power increased over the next decade, and with the use of data for machine learning , machine learning algorithms paved the way for deep learning.

These are Garry Kasparov's words : *"A weak human player plus a machine plus a better process is superior to a very powerful machine alone, but more remarkably, is superior to a strong human player plus machine and an inferior process."*

Deep learning is not symbolic AI . We have worked relentlessly to create hundreds of thousands of algorithms. Deep learning enables algorithms to write themselves. Given a dataset, they figure out for themselves. A human does not have to anticipate what would be the next algorithm or code to use to solve a problem.

Deepmind, a team of scientists, engineers and machine learning experts evolved with the idea of using AI for public benefit and scientific discovery , and collaborating with others on critical challenges , ensuring safety and ethics are highest priority. Deepmind's proposed approach was to dramatically accelerate the process of discovering new RL algorithms by automating the process of discovery in a data driven way. This resulted in a paradign shift from - from the deep blue to deep mind, from building a program to building an environment where the resulting algorithm is efficient.

But what if the RL algorithm captures unintended bias in the training set of environments. How do we remove unintended outcomes? This is the challenge with singularity.

The rate at which technology has progressed is is exponential. With singularity ,we may not be participants in the development of technology , self improving AI.

9. Robot rights. How do we define the humane treatment of AI?

We see robots doing the work of waiting tables, bell boys at hotel lobbies and greeting visitors. Robots are now the new species on earth. What is their identity?

Isaac Aasimov (1942) in his book “I robot” coined three laws for robots:

1) Robots must not harm humans or allow them to be harmed through inaction.

2) Robots must obey human commands unless doing so would violate the first rule.

3) Robots must protect themselves unless doing so would violate the first two rules.

Though these rules are fictional and from a long time back, they have helped innovators think about robot rights.

In some contexts, they make sense. Shouldn’t drones used in warfare follow them and make sure they don’t kill innocent people? Robot rights are important and must be developed as much as human rights have been.

Are we there yet? Are AI systems that we develop and deploy being evaluated with these commonsense laws in mind and how do we evaluate compliance?

Extending the concept to human rights to the new species, what are the basic rights a robot needs and do they need an identity?

10. Post privacy era. How do we define and protect privacy in the age of machine learning?

Every country has privacy laws. For example, https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/ governs how an individual’s personal information is used, collected or disclosed. Privacy laws are enacted at various levels,

Personal information can only be used for the purposes for which it was collected. If an organization is going to use it for another purpose, they must obtain consent again. Personal information must be protected by appropriate safeguards.

This safeguarded an individual’s privacy in the past. Today, every individual is continuously feeding data in to the network. Data collected on the cloud is not local , confined to a region and does not have boundaries.

Data collectors stand to gain from the process of collection. We need AI governance to determine how the gains (with regards to data) from AI are shared.

Telematics has been around for a long time. Drivers provided driving behaviour to insurance companies via a dongle connected to the car’s OBD port. Data was not collected without the driver’s consent. We now have the technology to enable collection of data directly from car manufacturers .Auto insurers can partner with a car manufacturer to get driver data and provide insurance coverage based on that information, It is a great idea for the insurer. The company can claim that products developed using the data will be more affordable.To the driver, the behind- the scenes agreements are unclear. Who owns the driver's data ? What rights does the owner have ?

Tesla offers in-built insurance. Volvo has a subscription service that includes insurance. When businesses gain data, they risk losing their user's trust without strong privacy protections.

Differentially private algorithms can be used to produce aggregate statistics over numeric data sets containing private or sensitive information.

Traditional Machine learning requires centralized training data. Cloud infrastructures have made it possible to centralize date. Federated learning though limited in scope allows for smarter models, lower latency, and less power consumption, all while ensuring privacy. Applying federated learning to mobile phone applications is a privacy feature. It requires machine learning practitioners to adopt new tools and a new way of thinking: model development, training, and evaluation with no direct access to or labeling of raw data, with communication cost as a limiting factor. It has limited scope. Google is implementing differential privacy algorithms and federated learning.

Post-privacy, AI and ML technologies should be capable of detecting and mitigating fraud, privacy breaches , threats in real time so that the data collected is not misused.