Co-Creating Innovation in Machine Learning - Intro

Motivation:

From ChatGPT to self-driving cars, "Intelligent Machines" powered by Machine Learning systems are increasingly helping us make decisions and predictions across a wide range of complex challenges. Recent studies, such as the McKinsey Global Survey on AI, show that AI adoption continues to grow steadily [1], with projections estimating the Artificial Intelligence market will reach 1,339.1 billion USD by 2030 [2]. Furthermore, advancements in research have led to innovative applications of Machine Learning algorithms beyond traditional business analytics.

This short project introduction offers a (somewhat) technical overview of the development process for Machine Learning systems, outlines key challenges faced by innovators, and presents examples of how Machine Learning algorithms have been used to create "Intelligent Machines."
Building on this foundation, I would like you to explore (or develop) innovative Machine Learning Systems (“Intelligent Machines”) in this project, while also identifying potential design challenges that developers encountered (or may encounter) during their design process.

Machine Learning - A Quick definition:

In its most general form, a Machine Learning Algorithms learns by analysing existing patterns in an available dataset to formulate a hypothesis (prediction / decision) for new data [3]. The algorithm formulates these hypotheses autonomously, i.e., without the need for explicit programming.
This puts Machine Learning in n unique interface between the following fields:

- **Mathematics** *(Study of quantity, structure, space and change)*
- **Statistics** *(Data analysis: from hypothesis to validation by data)*
- **Computer Science** *(Theory, experimentation and engineering for design and use of computers)*
- **AI & Machine Intelligence** *(Intelligent agents with perception and actions to achieve goals)*.

When innovators aim to create "Intelligent Machines" using Machine Learning algorithms, they typically follow a standardized, practical development process.

Designing “Intelligent Machines” using ML Algorithms:

This section provides a brief introduction to the design process of Machine Learning systems. These systems are typically developed through a cyclic process, which can be divided into three key stages: **Design**, **Data Collection**, and **Experimentation** (see Figure 1). At each stage, the designer or innovator must define critical system characteristics that often influence the subsequent stages of the development cycle. The process begins with the Design stage in each iteration.

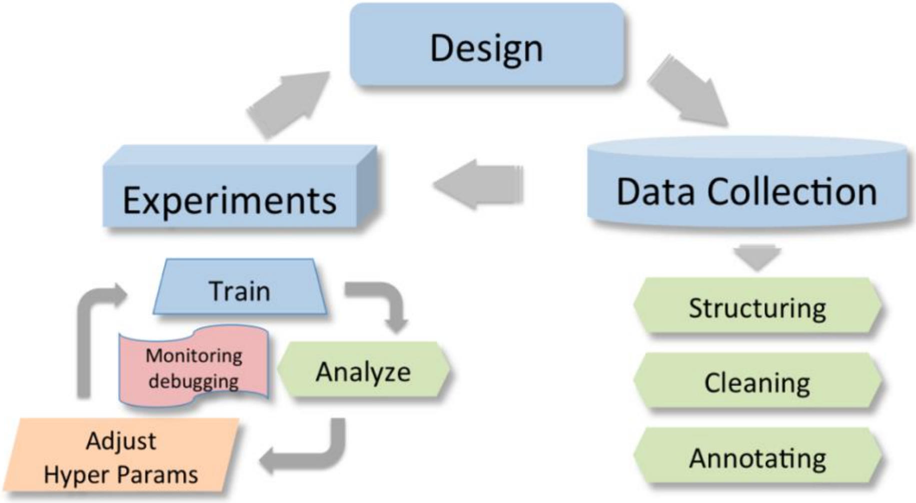


Figure 1: Practical Design Process for ML Applications [4]

Stage 1: Design

At the end of the Design stage you should have established a reasonable “End-to-End” system. For this purpose, you will need to answer some key design questions:

1. Was the task/ topic studied before?
    - Do literature review! Start from competitions/survey papers
2. What type of learning task must the ML algorithm complete?
    - **Common Types:** Classification, Regression, Clustering, Anomaly Detection, Dimensionality Reduction, Feature Selection
    - A short list of examples for each learning task can be found in Table 1 in the Appendix
    - Some learning tasks requrie specific ML model structure, training methods and data types
3. What type of Machine Learning model should I use?
    - Answers to this question typically require some previous experience and a deeper understanding of the issue
    - A brief overview of different Machine Learning models can be found in in the Appendix (see Table 1)

Important Consideration for the Design Stage:

If a model is making incorrect predictions, we are interested to know why and how to fix the model. Interpretability/explainability allows us to understand why a model makes a particular decision or prediction, for:

- Debugging a model.
- Extracting knowledge learned by a model.
- Creating trust between a model and its users.

Some types of Machine Learning Models (e.g. Deep Neural Networks), have relations between model parameters and input which can make it very difficult to interpret. [4]

Stage 2: Data Collection

At the end of the Data collection stage, you should have collected and pre-processed an adequate data set with which you can train your Machine Learning Model. Depending on the type of learning task, you might have to collect specific types of data and perform significant preprocessing.

Typically, we distinguish between two types of data: **Structured Data** and **Unstructured Data** (see Figure 2).

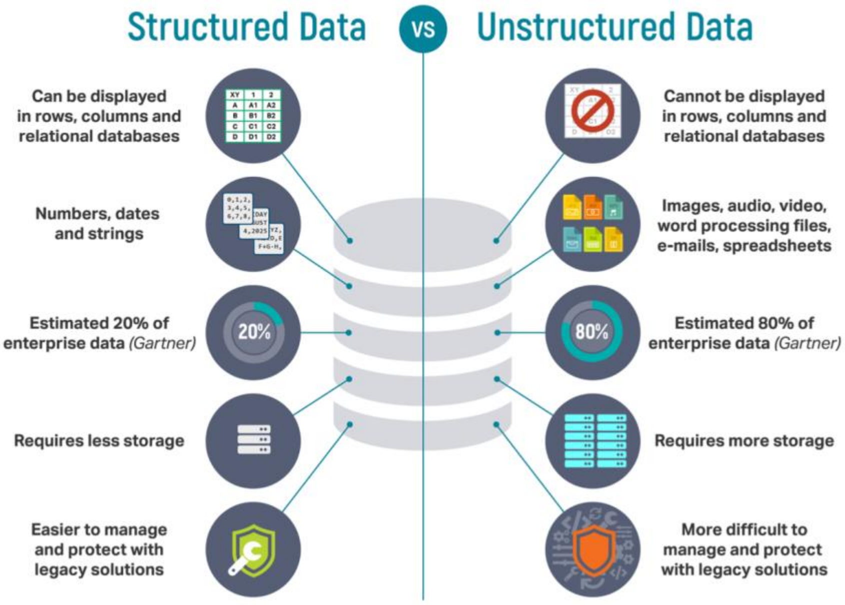
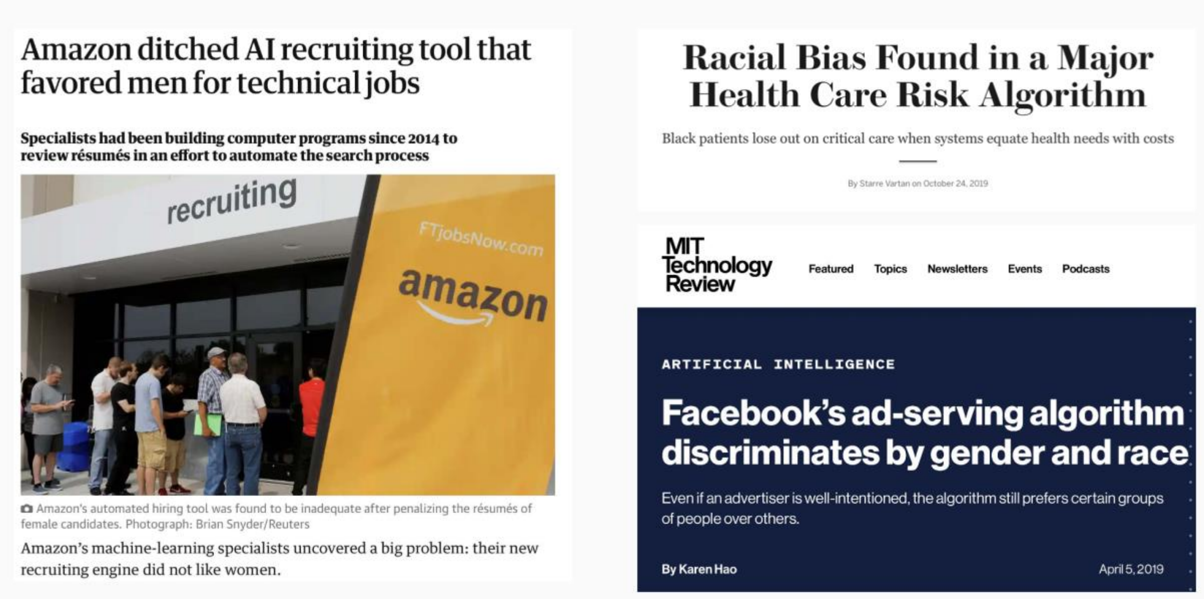

Structured Data
VS
Unstructured Data
Can be displayed in rows, columns and relational databases
Cannot be displayed in rows, columns and relational databases
Numbers, dates and strings
Images, audio, video, word processing files, e-mails, spreadsheets
Estimated 20% of enterprise data (Gartner)
20%
80%
Estimated 80% of enterprise data (Gartner)
Requires less storage
Requires more storage
Easier to manage and protect with legacy solutions
More difficult to manage and protect with legacy solutions

Figure 2 - Structured vs Unstructured Data [3]

Important considerations for the Data Collection Stage:

1. The Vapnik-Chevonenkis Inequality and the rule of 10
   - Each parameter in a Machine Learning Model requires at leat 10 data points to train properly
   - AlexNet (the first Convolutional Neural Network for image classification) has 62,378,344 parameters and was trained on a dataset of **1.2 million images**
     (if you want to find out more: https://en.wikipedia.org/wiki/AlexNet)

2. Data bias
   - Machine Learning models can often mirror pre-existing biases in our society.This is often arises when one-sided or non-representative datasets are collected. [5] (see Figure 3 for examples)
   - Bias can be mitigated by:
     i. Pre-processing (i.e., collecting data where sensitive attributes are independent from other features) [5]
     ii. In-processing (i.e., improving the learning algorithm)
     iii. Post-processing (i.e., adjusting a learned model to exclude sensitive attributes)
3. Fairness
   - If we cannot solve the bias issue, we can try to create a Fair algorithm. However this remains an open problem.
   - Example: The Leuphana University decides to introduce a Machine Learning Model to automate the University admission process. Do we require that:
     i. *The rate of acceptance should be independent of race/gender/belief*? (Independence)
     ii. *Among pople who deserve admission, the rate of acceptance should be independent of race/gender/belief?* (Separation)
     iii. *If the model accepts a particular student, the probability that the student deserves acceptance should be independent of their race/gender/belief?* (Sufficiency)



Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process

Amazon's automated hiring tool was found to be inadequate after penalizing the résumés of female candidates. Photograph: Brian Snyder/Reuters

Amazon's machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Racial Bias Found in a Major Health Care Risk Algorithm

Black patients lose out on critical care when systems equate health needs with costs

By Starre Vartan on October 24, 2019

ARTIFICIAL INTELLIGENCE

Facebook's ad-serving algorithm discriminates by gender and race

Even if an advertiser is well-intentioned, the algorithm still prefers certain groups of people over others.

By Karen Hao

April 5, 2019

Figure 3 - Consequences of Machine Learning Algorithms trained on unbiased data [5]

Stage 3: Experiment

At the end of the Experiment stage, your Machine Learning model should have learned to perform their assigned task using a representative dataset from your collected data. Moreover, you should have tested your ML model’s performance using predefined performance metrics (e.g., accuracy) and (depending on the learning task) a test dataset from your collected data which wasn’t used in the trainining of the model.

The appropriate learning method can depend on the learning taks, the chosen model and the collected data.

Typically, three different learning methods are used to train Machine Learning models: **Supervised learning** (see Figure 4)**,** **Unsupervised Learning** (see Figure 5) or **Reinforcement Learning** (see Figure 6).

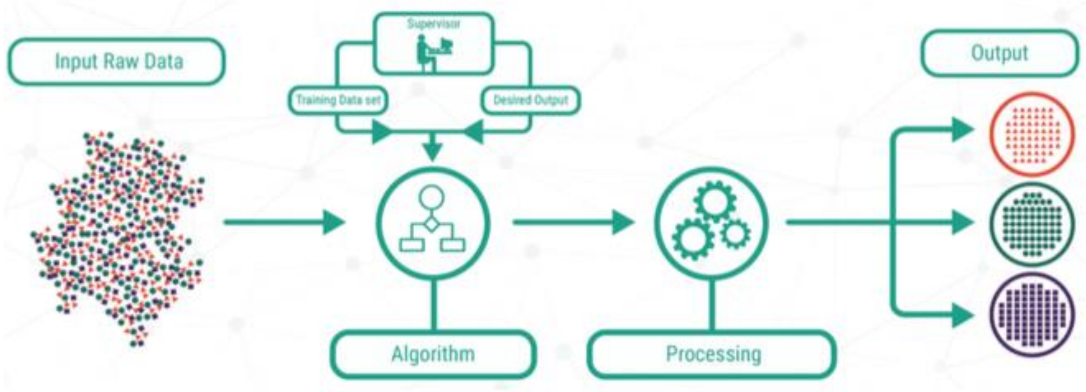


Figure 4 - Supervised Learning Method: The ML model learns excplicitly from data with clearly defined input-output pairs and direct supervisor feedback (Suitable for e.g., Classification, Regression or Ranking tasks) [3]

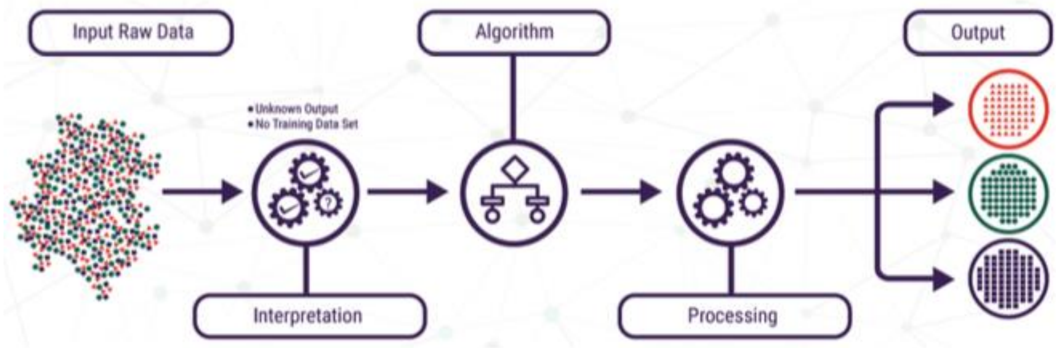


Figure 5 - Unsupervised Learning Method: The model attemps to understand the data and find patterns/ structures. Only the input data is given and no specific predictions are made. (Suitable tasks: Clustering, association mining, dimensionality reduction) [3]

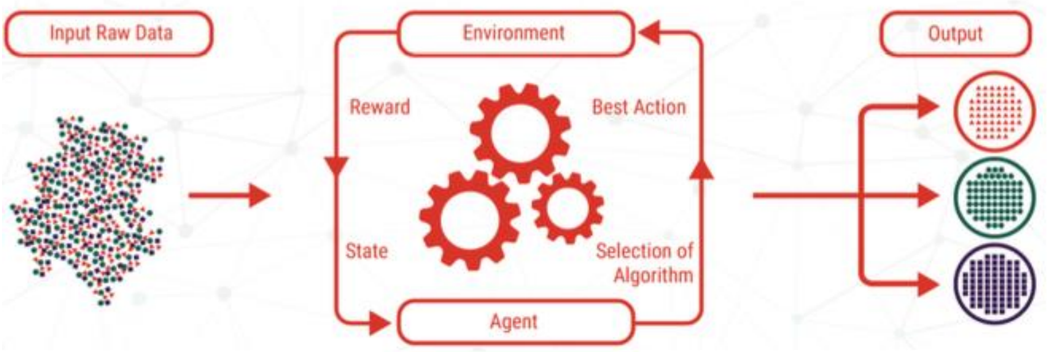
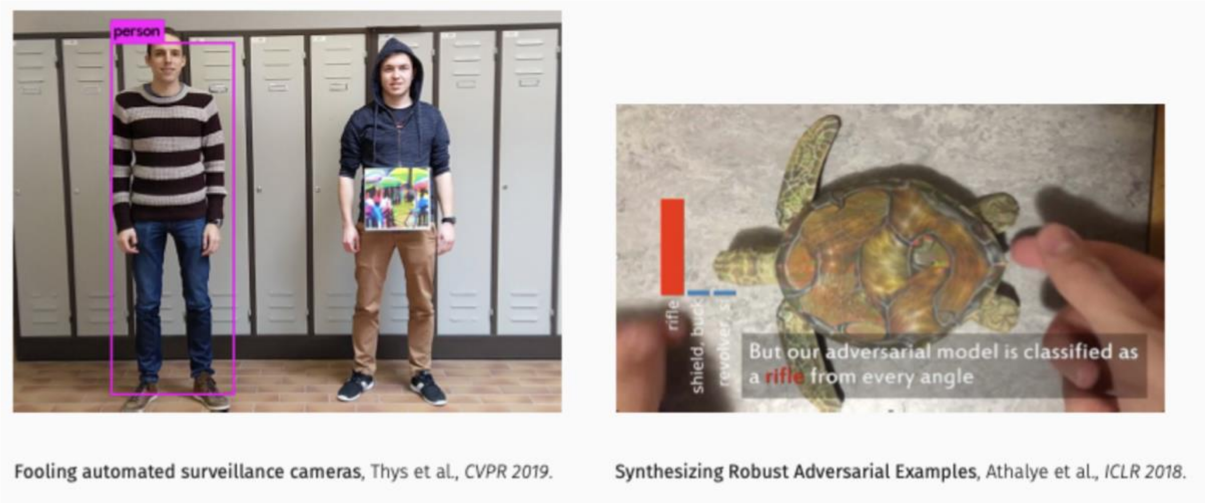


Figure 6 - Reinforcement Learning Method: Learns how to act in a given environment to maximise rewards. Only inputs, some output, some valuation and a decision process / reward process is given [3]

Important Consideration for the Experiment Stage:

Machine Learning models can be deployed in security- and safety-critical systems. When testing the trained model in the experiment stage, it is important to check its robustness to adversarial attacks (see Figure 7).


person
rifle
shield
revolver
But our adversarial model is classified as a rifle from every angle

Fooling automated surveillance cameras, Thys et al., *CVPR 2019*.

Synthesizing Robust Adversarial Examples, Athalye et al., *ICLR 2018*.

Figure 7 - Examples of adversarial attacks for Machine Learning algorithms used in image detection / classification

If you find that your model has some robustness issues during the Experiment stage, you can improve its robustness using different methods:

- Adversarial training (i.e., training your algorithm with data that it finds hard to process correctly)
- Data poisoning/ fuzzing (altering or augmenting training data to improve the models performance for data that slightly deviates from your collected dataset)
- Collecting a bigger dataset (which covers these edge cases)

Interestingly, Generative Adversarial Networks (a specific type of Machine Learning model) can help developers to generate new data for this purpose, based on their collected datasets [6].

Examples of modern applications of Machine Learning algorithms:

In this section, we will briefly explore some modern applications of Machine Learning algorithms that have led to the development of "Intelligent Machines."
Each example is accompanied by a short news article and an optional technical article for those interested in delving into the technical details. To help you assess your understanding, I’ve included a brief section with three questions for each news article.
Take your time reading the news articles and consider the decisions made during the Design, Data Collection, and Experimentation stages of the Machine Learning system discussed. For each example, I’ve highlighted interesting snippets from the articles and added comments that illustrate how we can identify potential design challenges.

Example 1: Medical Imaging using Machine Learning

“AI model identifies certain breast tumor stages likely to progress to invasive cancer
The model could help clinicians assess breast cancer stage and ultimately help in reducing overtreatment. ” [7]

Article:
https://news.mit.edu/2024/ai-model-identifies-certain-breast-tumor-stages-0722
Technical article (OPTIONAL!):
https://www.nature.com/articles/s41467-024-50285-1

Questions:

1. What existing issue in the field of DCIS detection does the use of Machine Learning solve?
2. What type of data does the Machine Learning model use to identify the different stages of DCIS?
3. What type of task does the Machine Learning model perform?

Answers:
1. Typically DCIS screening was an invasive procedure and very costly
2. Unstructured data (labelled images of breast tissue)
3. Clustering and Inference

Spotlight on a short Article snippet:

“First, they created a dataset containing 560 tissue sample images from 122 patients at three different stages of disease. They used this dataset to train an AI model that learns a representation of the state of each cell in a tissue sample image, which it uses to infer the stage of a patient’s cancer.” [7]

1. Is 560 data points enough for the training of the model?
   a. How many parameters can we reliably train? (See Rule of 10)
2. How were the patients chosen?
   a. Are there underlying biases in the choice for patients? (e.g., Age, ethnicity; See Bias and Fairness)
3. What are the consequences of a false positive vs a false negative?
   a. How do we reflect this in the training of our model?

Example 2: Intelligent Boarder Controls

“WE TESTED EUROPE’S NEW LIE DETECTOR FOR TRAVELERS — AND IMMEDIATELY TRIGGERED A FALSE POSITIVE
4.5 million euros have been pumped into the virtual policeman project meant to judge the honesty of travelers. An expert calls the technology “not credible.” “ [8]

Article:
https://theintercept.com/2019/07/26/europe-border-control-ai-lie-detector/
Related Video (OPTIONAL! in German):
https://www.youtube.com/watch?v=SY_xV-Raq3A
Technical article (OPTIONAL!):
https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/690706/EPRS_IDA(2021)690706_EN.pdf

Questions:

1. What existing issue in the field of border controls does the use of Machine Learning try to solve?
2. What type of data is collected to feed the Machine Learning algorithm?
3. With what accuracy does the system operate?

Answers:
1. Automating and thus streamlining EU border security by verifying key traveller information before arrival
2. Unstructured data (live images of facial expressions and mirco-gestures)
3. 75% accuracy for an unbalanced participant group w.r.t. ethnicity and gender

Spotlight on short article snippets:

“IBorderCtrl’s lie detection system was developed in England by researchers at Manchester Metropolitan University, who say that the technology can pick up on “micro gestures” a person makes while answering questions on their computer, analyzing their facial expressions, gaze, and posture.” [8]

1. How do we account for different communication habits in non-European countries?
2. What happens if the image quality is poorer than the training data leading to mis-classified micro-gestures? (Adversarial training)

“A study produced by the researchers in Manchester tested iBorderCtrl on 32 people and said that their results showed the system had 75 percent accuracy. The researchers noted, however, that their participant group was unbalanced in terms of ethnicity and gender, as there were fewer Asian or Arabic participants than white Europeans, and fewer women than men.” [8]

1. How many parameters can be tested using a test set of 32 people? (Rule of 10)
2. How do you ensure fairness and minimal bias with an unbalanced (i.e., biased) training dataset? (Robustness)
3. What happens if a traveller is mistakenly denied entry due to a wrong model analysis? Is this risk worth it? (Ethics)

Finishing off the Intro

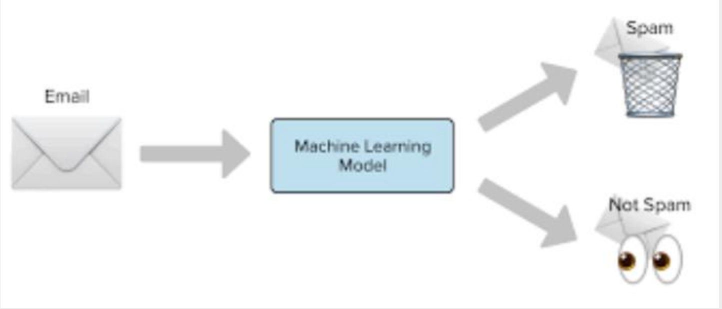
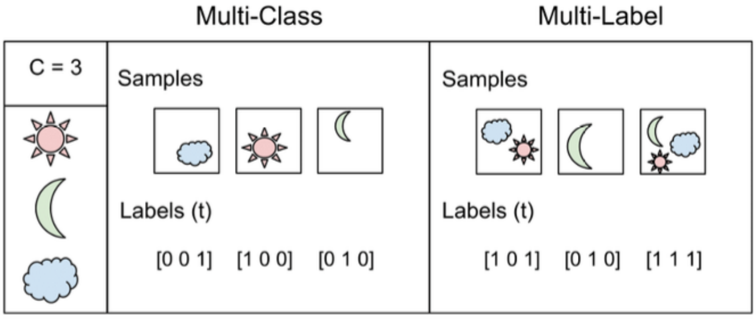
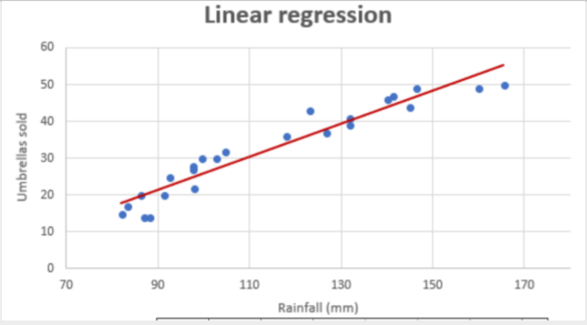
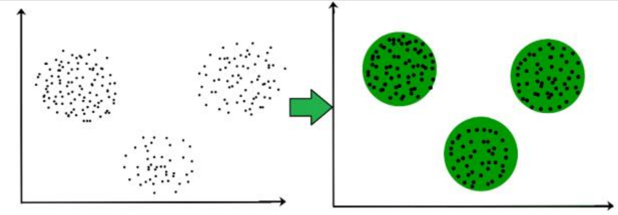
I have named this model "Co-creating Innovation in Machine Learning." Co-creation in innovation involves opening the innovation process to a broader range of voices that are often not heard. As AI and Machine Learning become more integrated into our daily lives, they are frequently still only developed by individuals who focus primarily on technical aspects.

While my brief introduction and your work during the welcome week won't make you an instant expert in the field, you will gain insights into fascinating applications and the key considerations and challenges involved in designing Intelligent Machines using Machine Learning algorithms.
I am confident that (over the course of this project) you will be able to come up with your own creative and innovative applications of Machine Learning to solve problems and critically analyse the buzzwords surrounding AI/ML that you often encounter in the media, thus actively Co-Creating Innovation in Machine Learning.

Appendix

Types of Learning tasks:

Machine Learning Algorithms are adapted to for specific types of learning tasks. The Table 1 shows key learning tasks and some examples for each task.

Learning Task	Example
Binary Classification	 [3]
Multiclass / Multilabel Classification	 Multi-Class Multi-Label C = 3 Samples Labels (t) [0 0 1] [1 0 0] [0 1 0] Samples Labels (t) [1 0 1] [0 1 0] [1 1 1] [3]
Regression (Univariate or Multivariate)	 [3]
Clustering	 [3]

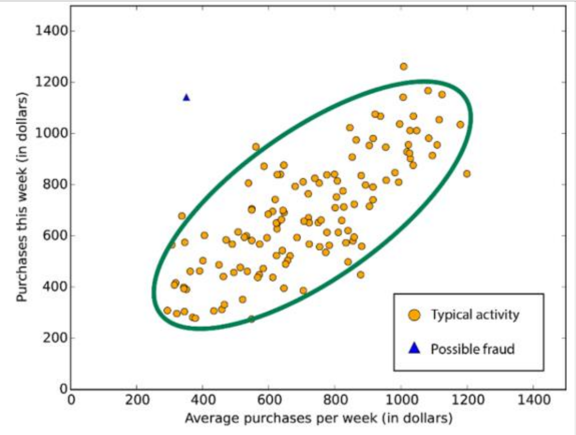
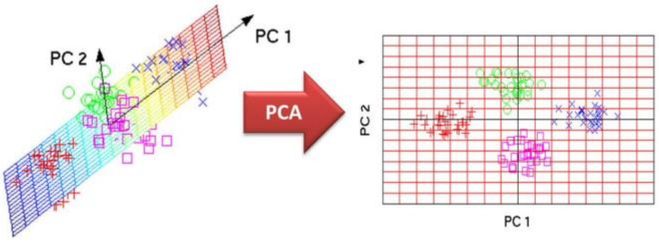
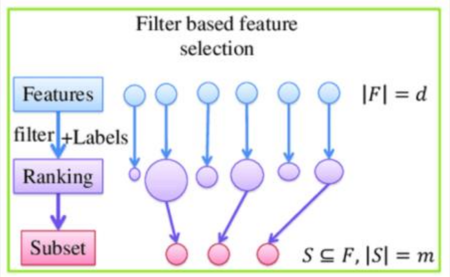
Anomaly Detection	 [3]
Dimensionality reduction	 [3]
Feature Selection	 Filter based feature selection Features filter +Labels Ranking Subset $\lvert F\rvert = d$ $S \subseteq F, \lvert S\rvert = m$ [3]

Table 1 - Learning Tasks and some Examples

Types of Machine Learning Algorithms:

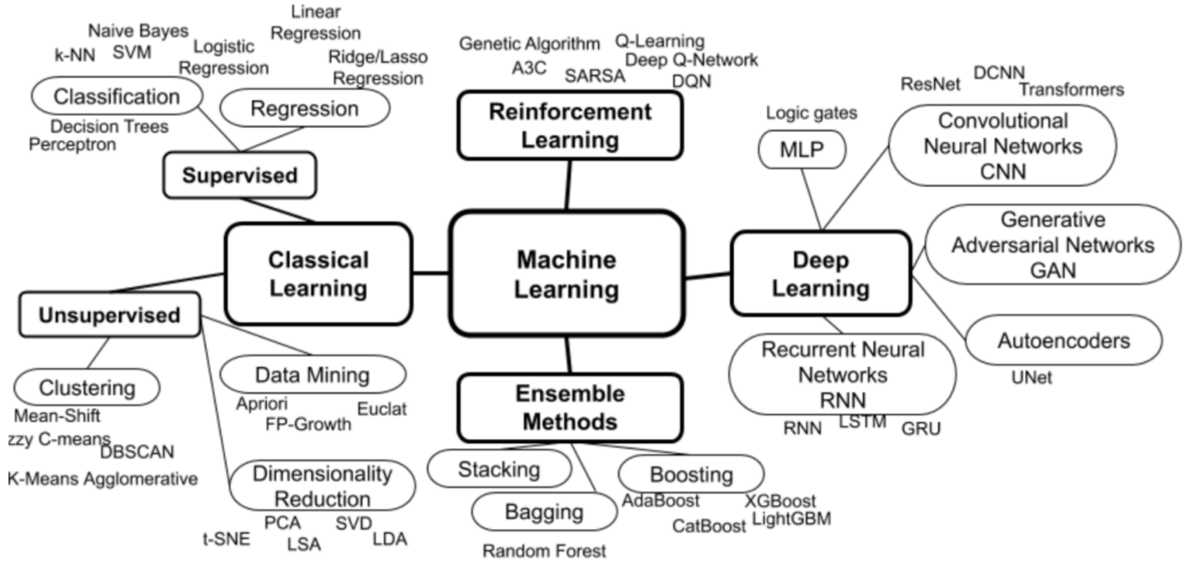


Figure 4 - Types of Machine Learning Algorithms (sorted by Learning Method) [4]

Bibliography

[1] M. Chui, B. Hall, A. Singla and A. Sukharevsky, “The state of AI in 2021,” 8 December 2021. [Online]. Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021. [Accessed 25 September 2024].

[2] Markets and Markets, “Artificial Intelligence (AI) Market,” May 2024. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-market-74851580.html. [Accessed 25 September 2024].

[3] K. Mikolajczyk and D. Gunduz, “ELEC60019 Machine Learning: Types of Learning,” Imperial College London, 2023. [Online]. Available: http://intranet.ee.ic.ac.uk/electricalengineering/eecourses_t4/course_content.asp?c=ELEC60019&s=J3.

[4] K. Mikolajczyk and C. Ciliberto, “ELEC60009 Deep Learning: Practical Development Process,” Imperial College London, 2023. [Online]. Available: http://intranet.ee.ic.ac.uk/electricalengineering/eecourses_t4/course_content.asp?c=ELEC60009&s=D3.

[5] S. Moosavi, “ELEC60009 Deep Learning: Reliability of Deep Learning,” 2023. [Online]. Available: http://intranet.ee.ic.ac.uk/electricalengineering/eecourses_t4/course_content.asp?c=ELEC60009&s=D3..

[6] V. Sandfort, K. Yan, P. J. Pickhard and R. M. Summers, “Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks,” 15 November 2019. [Online]. Available: https://www.nature.com/articles/s41598-019-52737-x. [Accessed 15 September 2024].

[7] A. Zewe, “AI model identifies certain breast tumor stages likely to progress to invasive cancer,” MIT News, 22 July 2024. [Online]. Available: https://news.mit.edu/2024/ai-model-identifies-certain-breast-tumor-stages-0722. [Accessed 25 September 2024].

[8] R. Gallagher and L. Jona, “WE TESTED EUROPE’S NEW LIE DETECTOR FOR TRAVELERS — AND IMMEDIATELY TRIGGERED A FALSE POSITIVE,” The Intercept_, 26 July 2019. [Online]. Available: https://theintercept.com/2019/07/26/europe-border-control-ai-lie-detector/. [Accessed 25 September 2024].