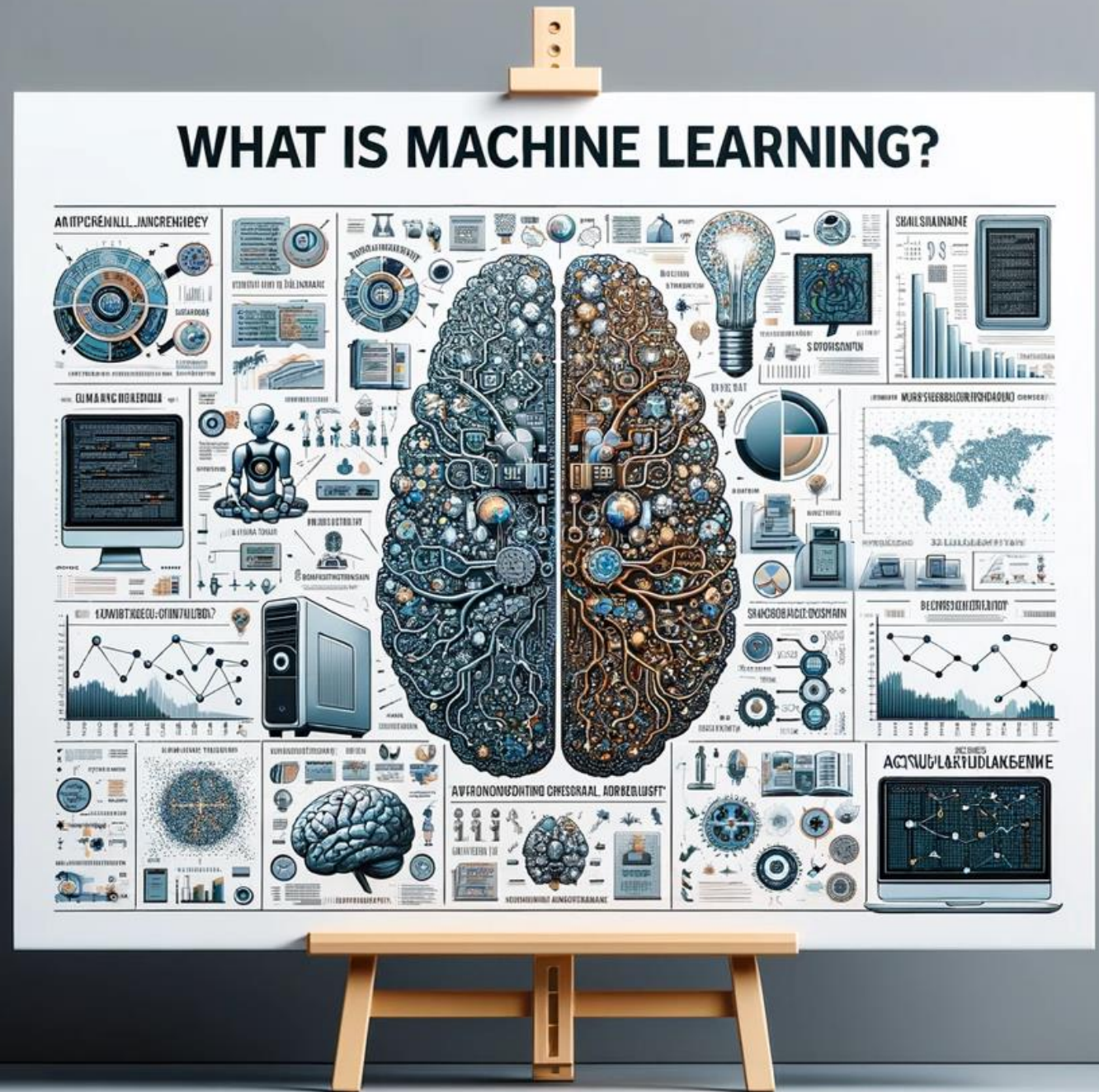


Prof. Dr. Mostafa Elhosseini

Professor of Smart Systems Engineering

<https://youtube.com/drmelhosseini>



| Agenda

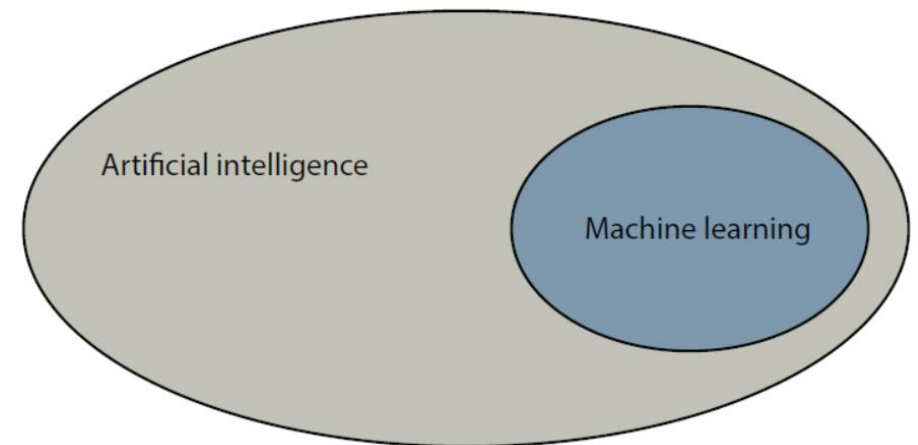
- Artificial Intelligence
- What is Machine Learning
- Spam Filter
- ML Applications
 - Disease Diagnosis System
 - Customer Churn Prediction
 - Fraud Detection System
 - Autonomous Vehicle Navigation
- ML Jargons

| Artificial Intelligence

- Artificial Intelligence (AI) refers to the development of computer systems that can perform tasks that typically require human intelligence.
 - Learning from experience (machine learning),
 - Understanding natural language,
 - Recognizing patterns,
 - Solving problems, and
 - Making decisions.

| What is Machine Learning

- Machine Learning is the science (and art) of programming computers so they can learn from data
 - Make decisions based on Data
- A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E . Tom Mitchell, 1997



| Inspiration

- Human learn by reading books, observing situations, studying in schools, exchanging conversations, browsing websites, among other means
- The algorithm learns from data, similar to how humans learn from experiences.

- A spam filter is a software tool that is used to detect unsolicited and unwanted email and prevent those messages from getting to a user's inbox

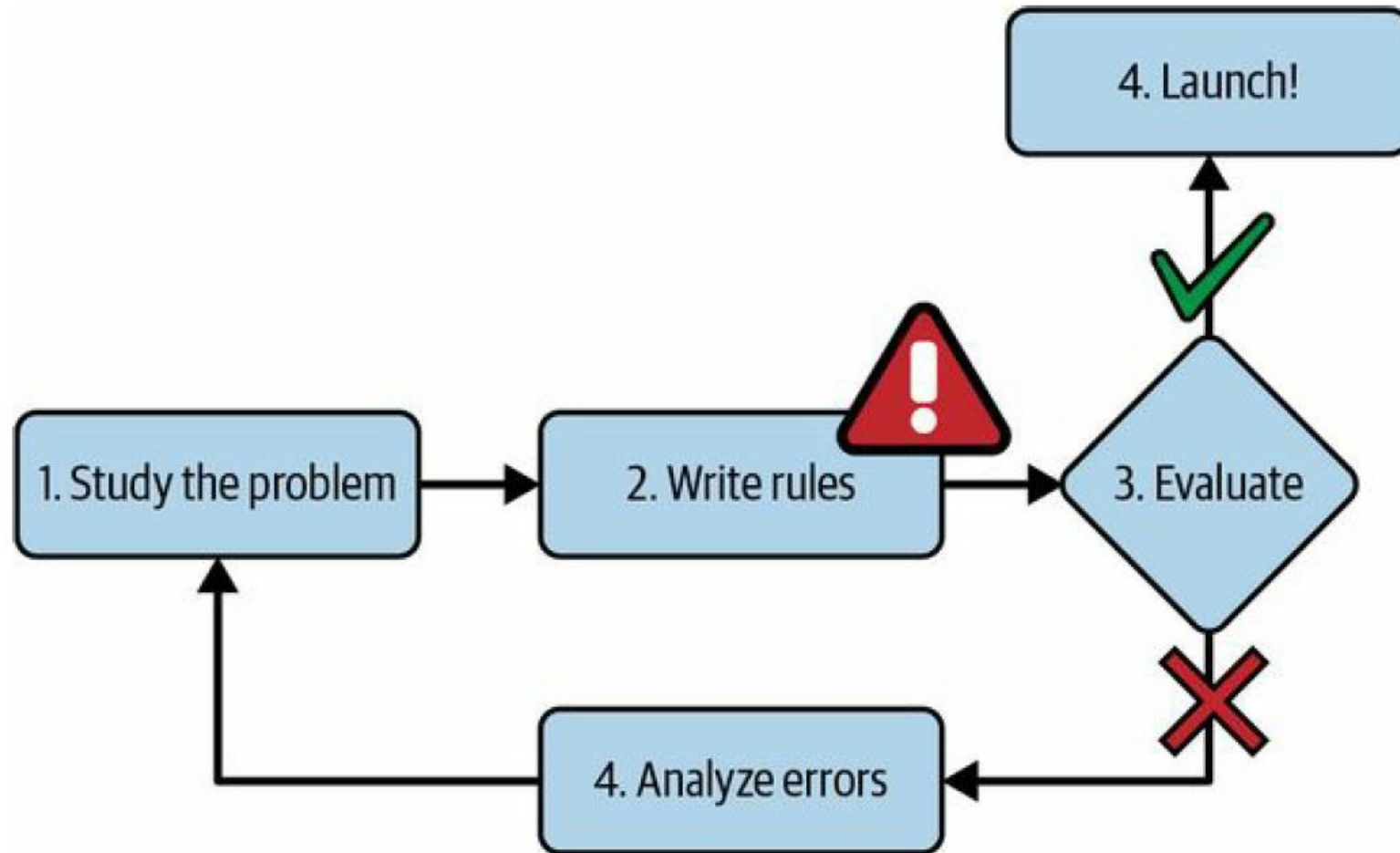


| Spam Filter

Traditional approach

- You might notice that some words or phrases (such as “4U,” “credit card,” “free,” and “amazing”) tend to come up a lot in the subject.
- Perhaps you would also notice a few other patterns in the sender’s name, the email’s body, and so on.
- You would write a detection algorithm for each of the patterns that you noticed, and your program would flag emails as spam if a number of these patterns are detected
 - Long list of complex rules

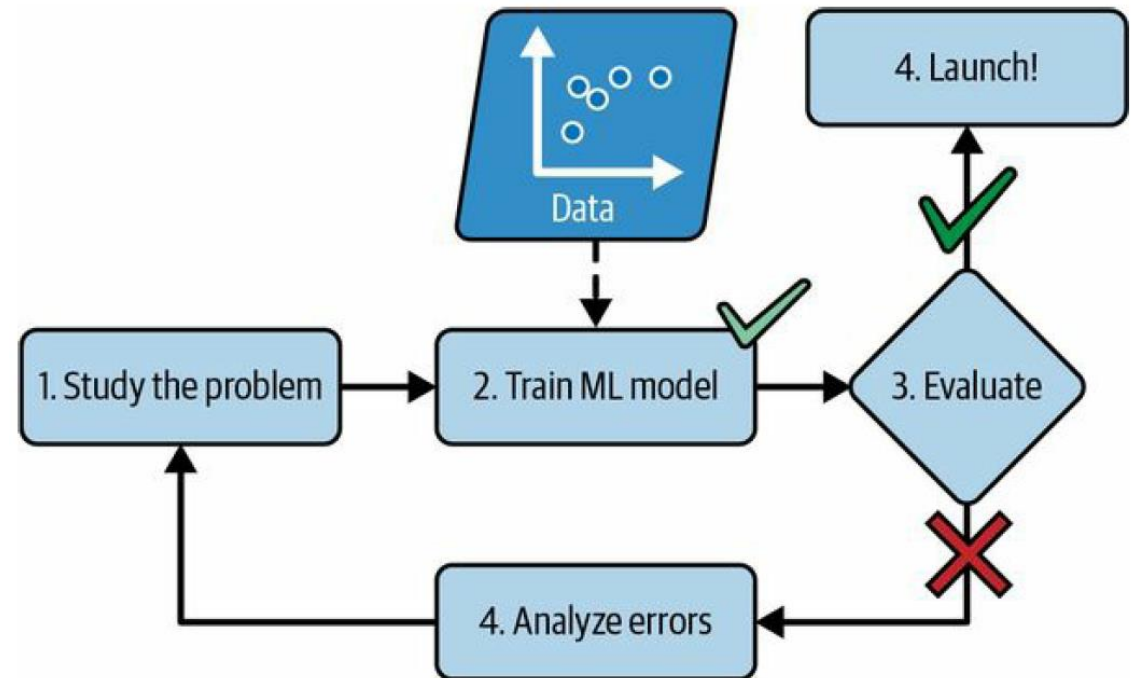
| Spam Filter



| Spam Filter

Machine Learning Approach

- **Automatically learns** which words and phrases are good predictors of spam by detecting unusually frequent patterns of words in the **spam** examples compared to the **ham** examples
- The program is much shorter, easier to maintain, and most likely more accurate



| Spam Filter

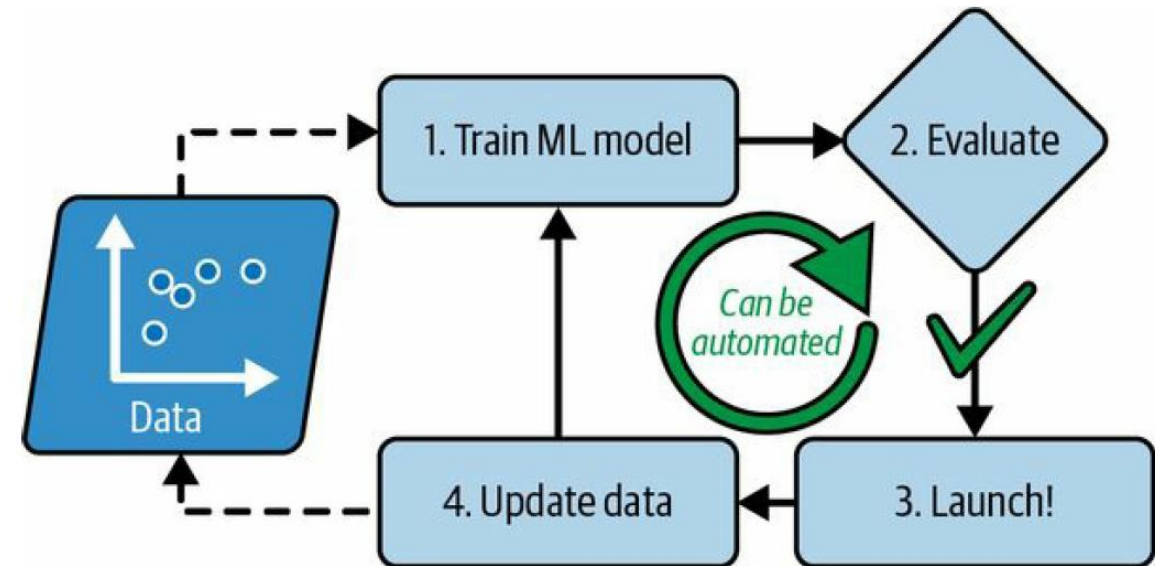
Traditional approach

- If spammers notice that all their emails containing “4U” are blocked, they might start writing “For U” instead.
- A spam filter using traditional programming techniques would need to be updated to flag “For U” emails.
- If spammers keep working around your spam filter, you will need to keep writing new rules forever.

| Spam Filter

Machine Learning Approach

- Spam filter based on Machine Learning techniques automatically notices that “For U” has become unusually frequent in spam flagged by users, and it starts flagging them without your intervention



| Spam Filter

- Your spam filter is a machine learning program that, given examples of **spam** emails (flagged by users) and examples of regular emails (nospam, also called “**ham**”), can learn to flag spam.
- The examples that the system uses to learn are called the **training set**.
- Each training example is called a **training instance** (or sample).
- The part of a machine learning system that learns and makes predictions is called a **model**.
- Neural networks and random forests are examples of models.

| Spam Filter

- A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E . Tom Mitchell, 1997.
- In this case, the task T is to flag spam for new emails, the experience E is the training data, and the performance measure P needs to be defined; for example, you can use the ratio of correctly classified emails. This particular performance measure is called accuracy, and it is often used in classification tasks.

| Disease Diagnosis System

- Task (T): Diagnosing diseases from medical imaging
- Performance Metric (P): Accuracy of diagnosis (percentage of correct diagnoses)
- Experience (E): Number and variety of medical images with confirmed diagnoses used for training



| Customer Churn Prediction

- Task (T): Predicting which customers are likely to stop using services/products
- Performance Metric (P): Precision and recall of churn predictions
- Experience (E): Historical data on customer interactions, purchases, and retention



| Fraud Detection System

- Task (T): Identifying fraudulent credit card transactions
- Performance Metric (P): The ratio of frauds correctly identified to the total transactions processed
- Experience (E): Datasets of transaction histories with labeled fraudulent and non-fraudulent transactions



| Autonomous Vehicle Navigation

- Task (T): Safe and efficient navigation of a vehicle in an urban environment
- Performance Metric (P): Number of successful trips without human intervention
- Experience (E): Accumulated miles driven and variety of traffic scenarios encountered



| ML Jargons

Car	Milage (in miles)	Age (in years)	Car Brand	Body State	Estimated Price (\$)	
Car 01	50,000	3	Toyota	Good	15,000	
Car 02	80,000	5	Honda	Excellent	12,000	
Car 03	30,000	2	Ford	Fair	20,000	
Car 04	100,000	7	Chevrolet	Poor	8,000	
Car 05	60,000	4	BMW	Excellent	18,000	
Car 06	45,000	3	Mercedes	Good	22,000	
Car 07	75,000	6	Nissan	Fair	13,000	
Car 08	55,000	4	Volkswagen	Good	16,000	
Car 09	90,000	8	Subaru	Poor	9,000	
Car 10	40,000	2	Hyundai	Excellent	19,000	

| ML Jargons

- Training set: the examples that the system uses to learn
- Training instance: Training sample -> each training example
- Label: spam or not
- Feature = attribute + value
- Predictors: features
- Ex. Price of a car
 - Set of features (mileage, age, brand) called predictors
 - Feature : Mileage = 15,000
- Performance measure P:
- Ratio of correctly classified emails.