

CSDS503 / COMP552 – Advanced Machine Learning

Faizad Ullah

About Me

□ Faizad Ullah

□ Ph.D. Student at LUMS

□ **Specialization**

- Natural Language Processing (NLP)
- Machine Learning
- Data Science

□ **Contributions**

- Medical Image Analysis
- Graph Analysis
- Text Analytics of Low-Resourced Language



Faizad Ullah

FOLLOW

Research Associate, [Lahore University of Management Sciences \(LUMS\)](#), Pakistan
Verified email at lums.edu.pk

Natural Language Processing Medical Image Analysis Deep Learning Data Science

GET MY OWN PROFILE

TITLE	CITED BY	YEAR
Brain MR image enhancement for tumor segmentation using 3D U-Net F Ullah, SU Ansari, M Hanif, MA Ayari, MEH Chowdhury, AA Khandakar, ... Sensors 21 (22), 7528	26	2021
Comparing prompt-based and standard fine-tuning for Urdu text classification F Ullah, U Azam, A Faheem, F Kamiran, A Karim Findings of the Association for Computational Linguistics: EMNLP 2023, 6747-6754	3	2023
Detecting Cybercrimes in Accordance with Pakistani Law: Dataset and Evaluation Using PLMs F Ullah, A Faheem, U Azam, MS Ayub, F Kamiran, A Karim Proceedings of the 2024 Joint International Conference on Computational ...		2024
UrduMASD: A Multimodal Abstractive Summarization Dataset for Urdu A Faheem, F Ullah, MS Ayub, A Karim Proceedings of the 2024 Joint International Conference on Computational ...		2024





Grading

Class Participation and Attendance	05%
Quizzes (5-8)	20%
Assignments (3-5)	10%
Midterm	25%
Final Term	30%
Project (Or Additional Assignments)	10%

Programming Tasks

- ❑ *3-5 Assignments
 - Programming Assignments
- ❑ *One Project
- ❑ Programming Environment
 - Python (Pytoch, TensorFlow, Colab)

*Vivas will be conducted for assignments and the project

Policies

☐ Quizzes

- Most quizzes are announced (50% quizzes will be unannounced)
- Announcements will be made during the class and on slides, so check slides regularly if you miss lectures. No announcement will be sent via email

☐ Sharing

- Copying is not allowed for assignments. Discussions are encouraged; however, you must submit your own work.
- Violators would be reported to the Disciplinary Committee or face marks reduction penalties

☐ Plagiarism

- Do NOT pass someone else's work as your own!
- Write in your own words and cite the reference if you use someone else's material.

Policies (2)

❑ Submission Policy

- Submissions are due at the day and time specified
- Late submissions will result in **10% marks deduction per day** from obtained marks.

❑ Attendance Policy

- You are advised to attend all lectures.
- It's the students' responsibility to recover any information or announcements posted during a lecture from which they were absent.

❑ Classroom behavior

- Maintain classroom sanctity by remaining **quiet and attentive**
- Asking questions is encouraged.
- You are not allowed to use a **Laptop/mobile phone**, etc., during class.

Policies (3)

☐ Retakes

- No retakes for quizzes, assignments, exams, or projects
- In case of any medical emergency or unavoidable circumstances, inform before hand and seek a formal approval. You need to share medical reports for departmental record.
- **Do not wait for the final exam to seek approval for retakes**

Course Material

☐ All course material (i.e., Books, class handouts, reading assignments) will be shared on Moodle

☐ Text Book

- Machine Learning: A Probabilistic Perspective, Murphy, Kevin P. MIT press, 2012 – Murphy.
- The Elements of Statistical Learning: Data mining, Inference, and Prediction, Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, Springer Science & Business Media, 2009 – ESLII

☐ Reference Book

- Machine Learning, Tom Mitchell, McGraw Hill, 1997 – TM

Contact

□ How to contact me?

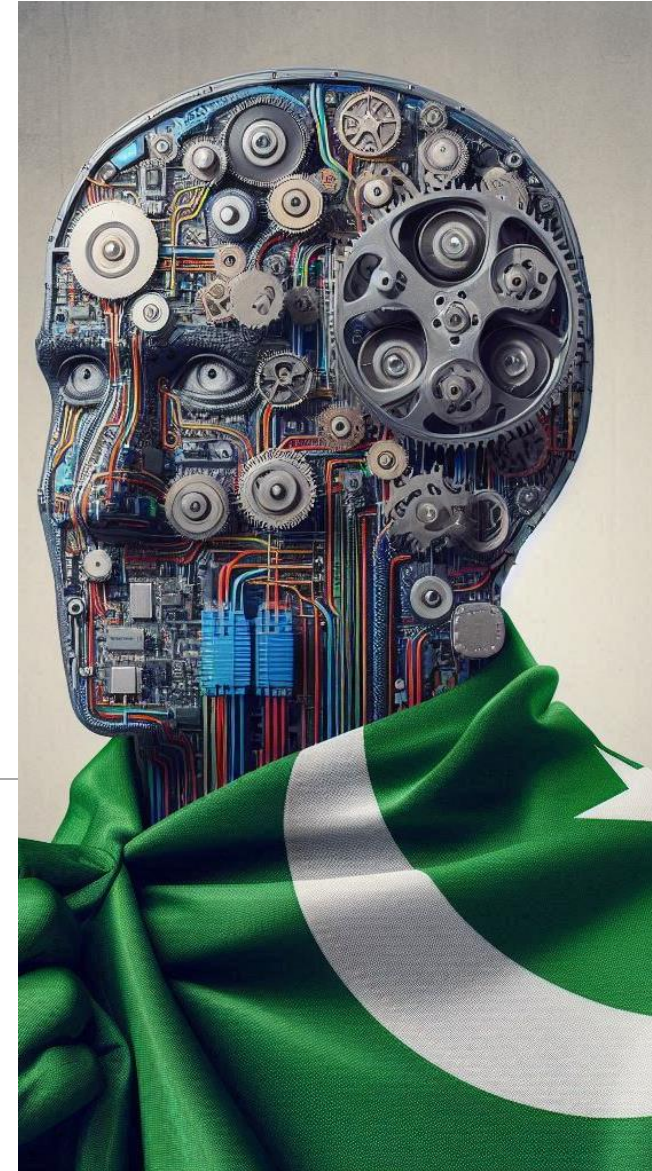
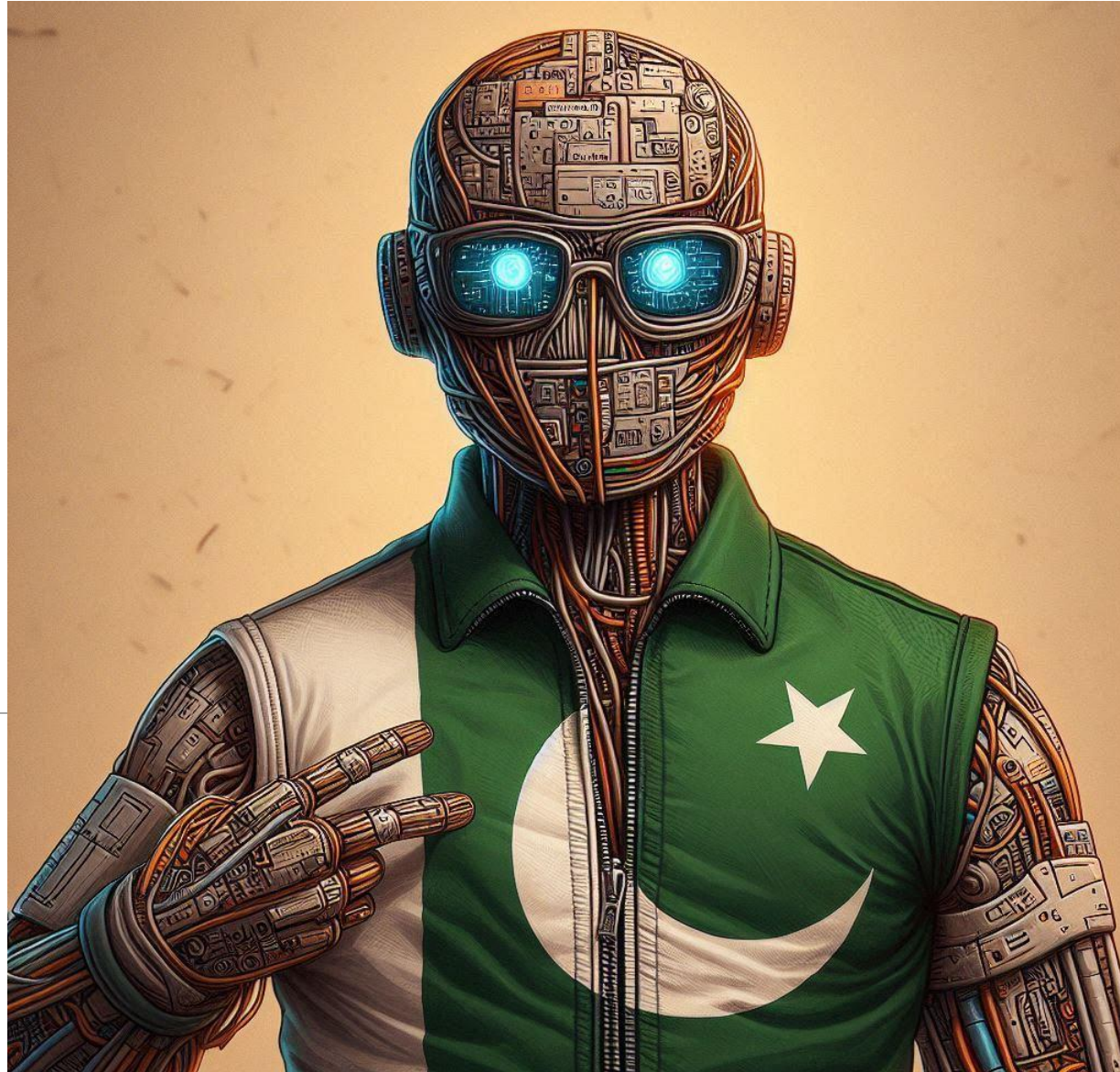
- E-mail: faizadullah@fccollege.edu.pk
- Office: 426-G
- Office Hours: Mentioned on office door

Most Important

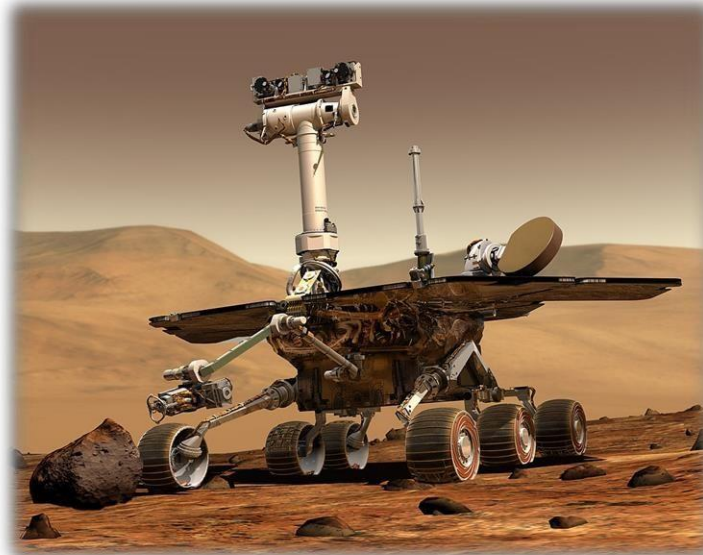
Don't be afraid of giving wrong answers!

Let's start our ML journey...

Machine Learning?

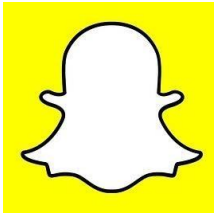


We Imagine Machine Learning as...



ML is all around us...

Google Translate



You Tube



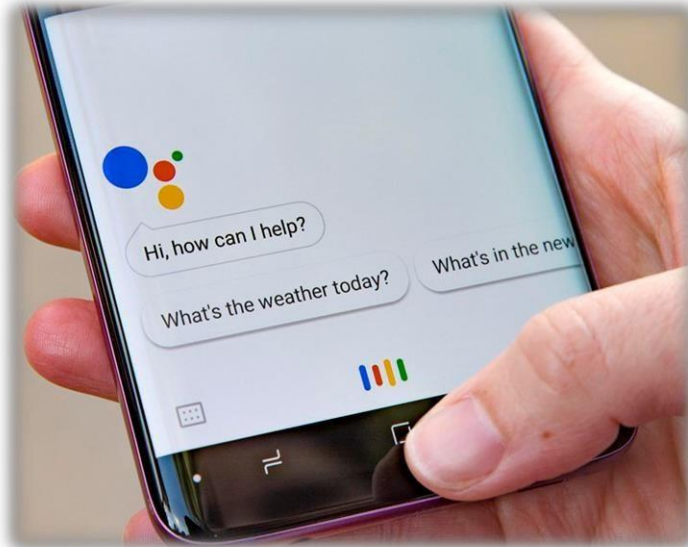
amazon alexa



Robots???



Robots we
Imagine



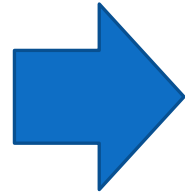
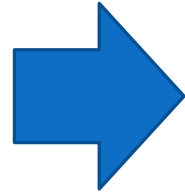
Actual Robots around us

Robots Invasion We Imagine



Robots Invasion We Imagine





Actual Invasion We Imagine

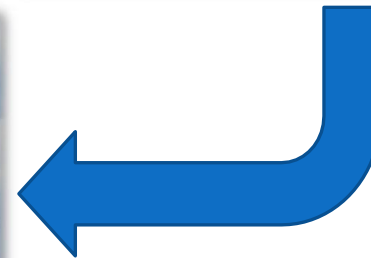
Why Discussing All This in ML Course?

- ❑ A Broader understanding of Machine Learning
- ❑ Although the focus of this course is concepts, mathematics, and implementation of machine learning algorithms.
 - But you should know why we needed ML
 - What comes after we have learned ML
 - How are ML algorithms deployed in real-world applications

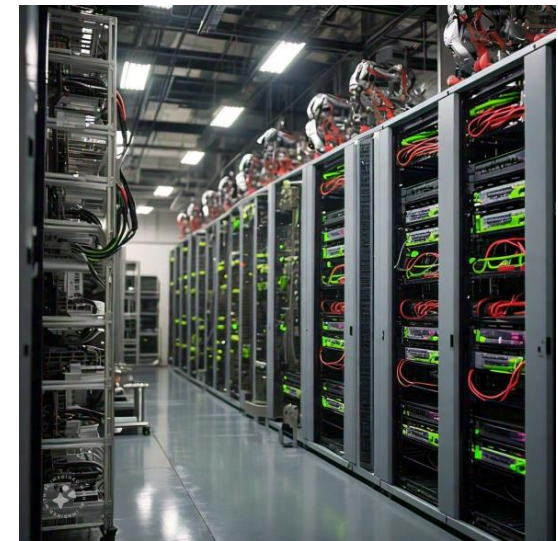
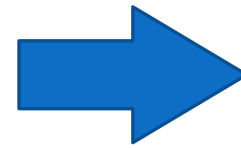
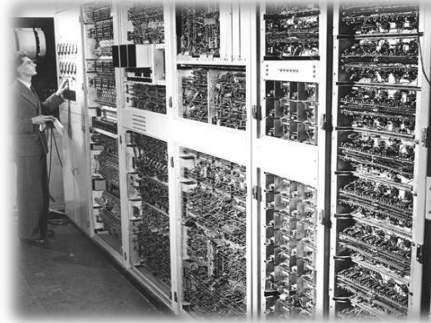
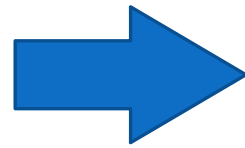
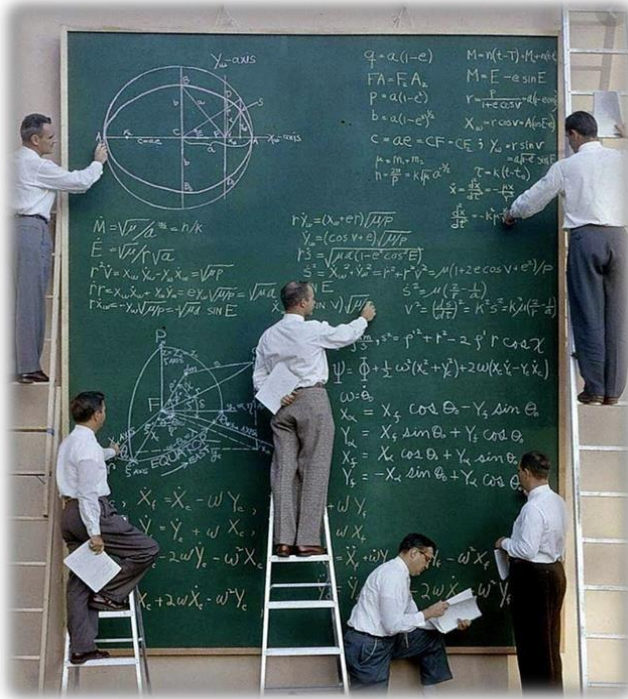
What is Machine Learning?

How does it work?

Machine as Mechanical Helpers



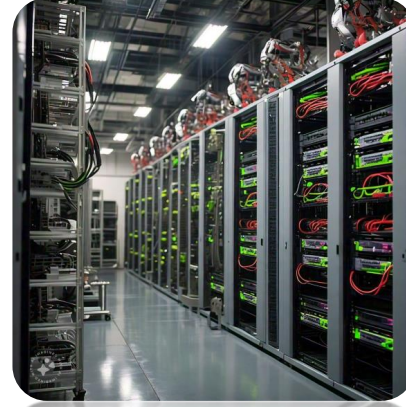
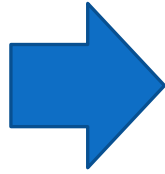
Machines as Intellectual Helpers



Machines as Intellectual Helpers



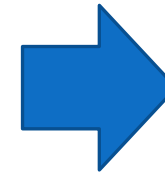
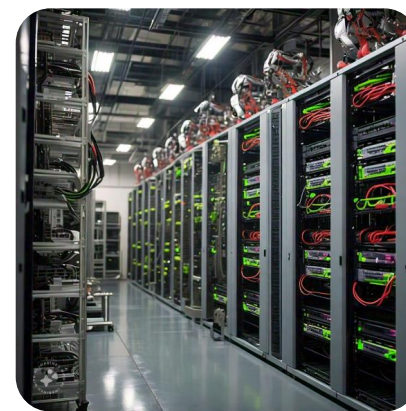
Is this a cat or a dog?



Dog



Should I hire this person?

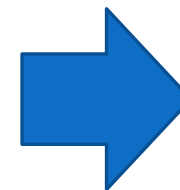
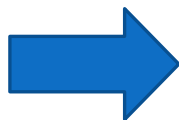


No

A Classifier

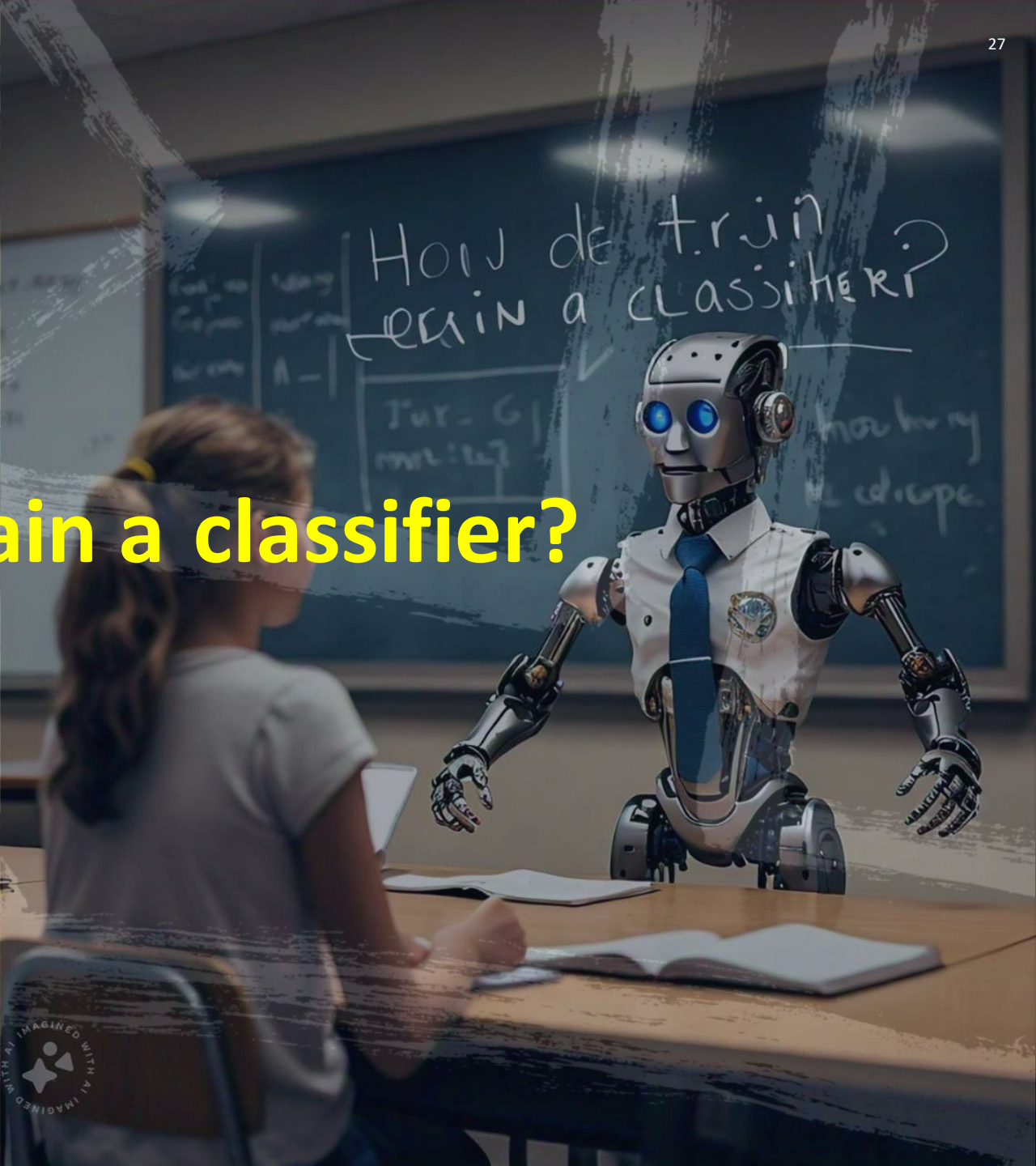


{happy, sad, angry,
surprised, neutral}



Happy,
surprised

How do we train a classifier?



How to train your intern?

☐ How would you train a new intern to conduct job interviews?

☐ Option 1: Teach all the complicated rules

- Grades are important
- University is important
- Great grades + Good university = All good!
- Bad grades + Unknown university = Not so good
- Bad grades + Good university = ?
- Good grades + Unknown university = ?

☐ Still there would be exceptions

☐ Very hard to instill intuitive and experiential knowledge



How to train your intern?

❑ **How would you train a new intern to conduct job interviews?**

❑ **Option 2:** Make them sit and watch, as a expert conduct interviews.

- Learning by experience
- Eventually, patterns start emerging
- Let the intern get the intuition on their own!

❑ More experience = Better learning

❑ More exposure (**balanced cases**) = Better learning

How to train your intern?

Caveat!

- What if the expert has systematic flaws of judgement aka biases?
 - Conduct sessions with many experts
 - What is they all share biases and stereotypes?
 - Initially, your intern could only be as good as the expert

How to train your machine?

- Allowing the machines *to learn on their own*, using *prior decisions of experts* is known as Machine Learning!

Supervised

The outcome is provided along with the data.

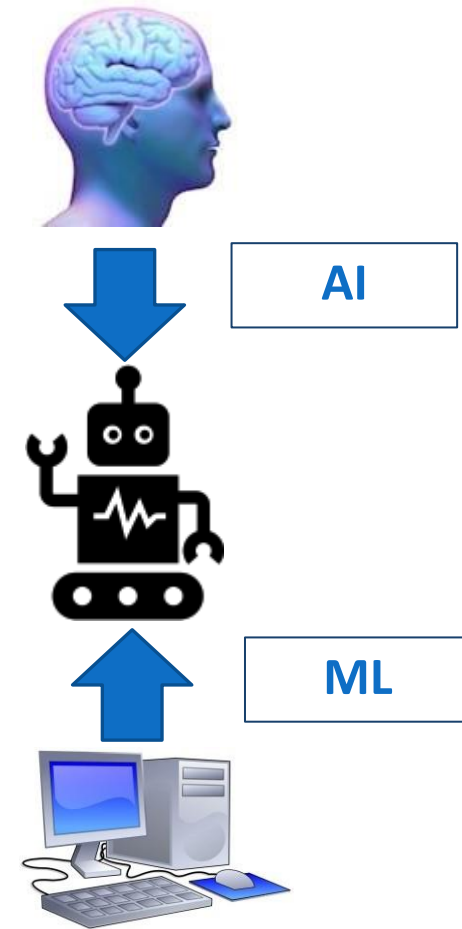
Unsupervised

The outcome is NOT provided along with the data.

Classifying emails in Spam and Not_Spam

Artificial Intelligence VS ML

- ❑ Colloquially both terms are used interchangeably
- ❑ However, traditionally there is a difference.
 - The goal of AI was to make a machine more like a human
 - Give the machine a lot of world knowledge
 - A logical decision-making framework
- ❑ The ML framework seeks to make a better machine – not necessarily emulating a human
 - Based on Statistics and Optimization – not logic!
 - Learn from labelled data
 - More data – more consistent decisions
 - More balanced data – more confident decisions



Life is not governed by certainty...

- ❑ **Certainty** in the real-world is a rare luxury
 - Probability of something of being 0 or 1 is very rare!
- ❑ **Uncertainty** is the basis of the ML that is quantified using probability and statistics
 - Something can and cannot happen with a certain probability!

Traditional Computer Science

□ Tasks like:

- Play an audio/video file
- Display a text file on screen
- Perform a mathematical operation on two numbers
- Sort an array of numbers using *Insertion Sort*
- Search for a string in a text file
- ...

Data



Program



Output

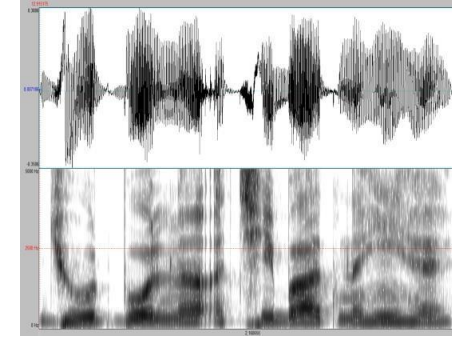
Problems that Traditional CS Can't Handle



Tumor? Y/N



Price?



What was said?

[illegible]

Summarize text

Data

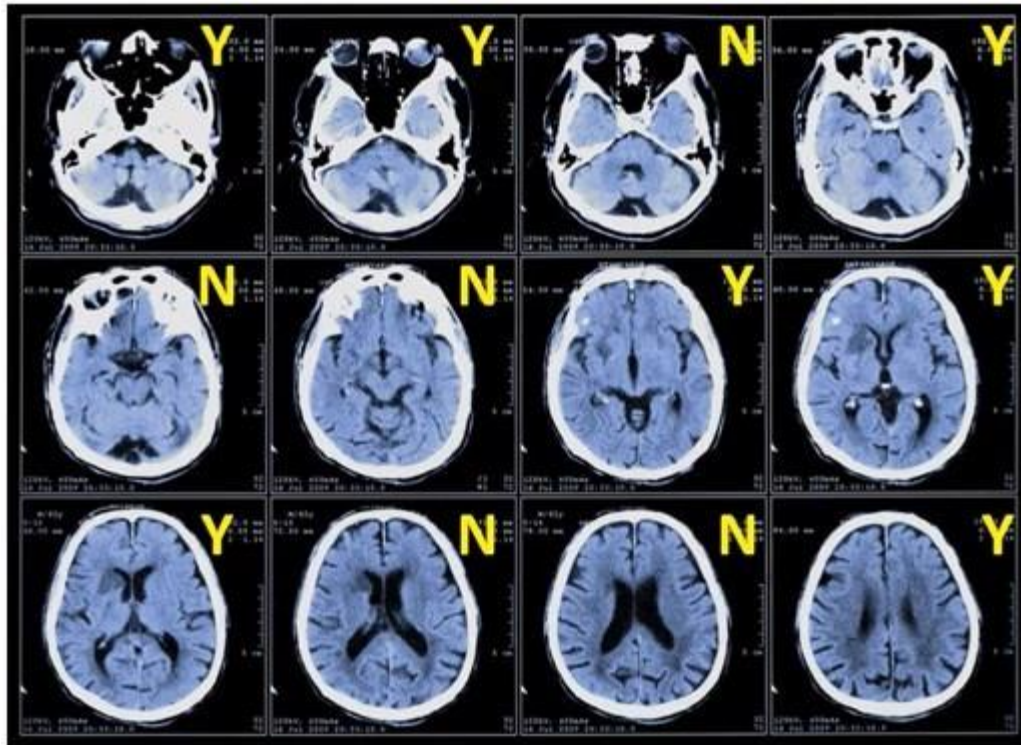
Program?



Output

Machine Learning

Classification



Regression



\$100,000



\$140,000



\$400,000



\$250,000



\$190,000

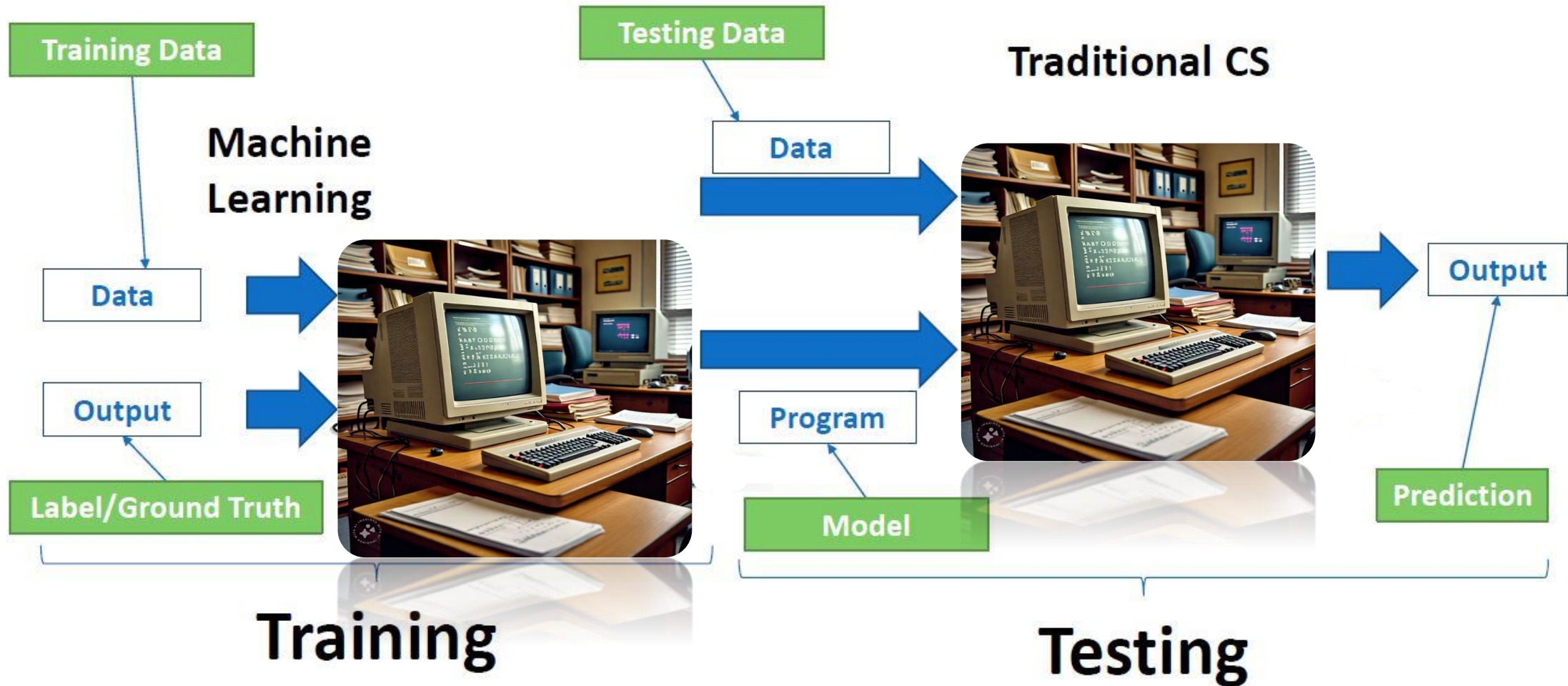
Traditional CS



Machine Learning



Machine Learning Pipeline



What is Machine Learning?

□ Formally:

- A computer program A is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E . (Tom Mitchell, 1997)

□ Informally:

- Algorithms that improve on some task with experience.

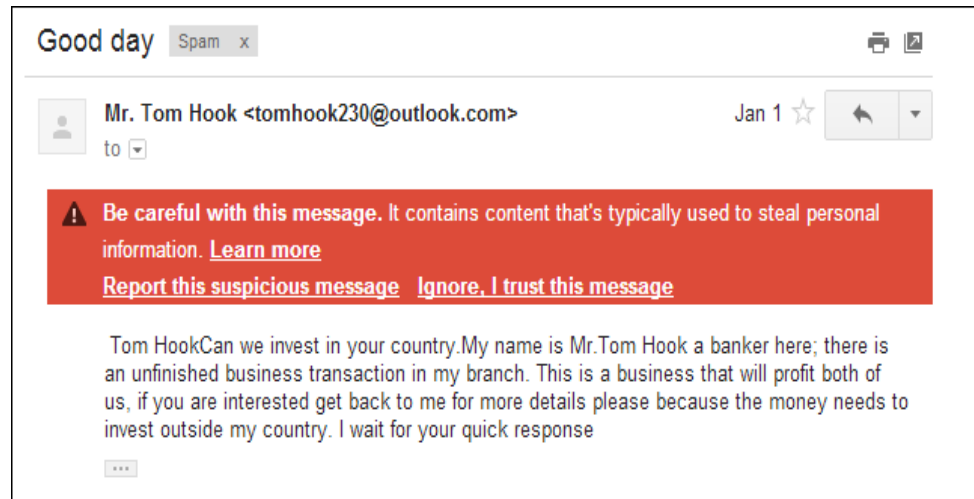
To train a classifier, we need labelled data (called dataset)

Data – Big, Big,... data!

□ How do we obtain these massive datasets to train our Machine Learning models?

- From real interactions e.g., call centers
- Expert annotators e.g., hired teams of annotators
- Crowd sourcing

Tagging



What makes you human?

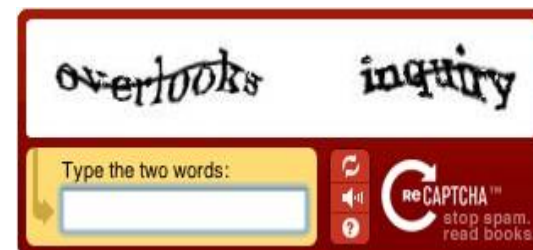
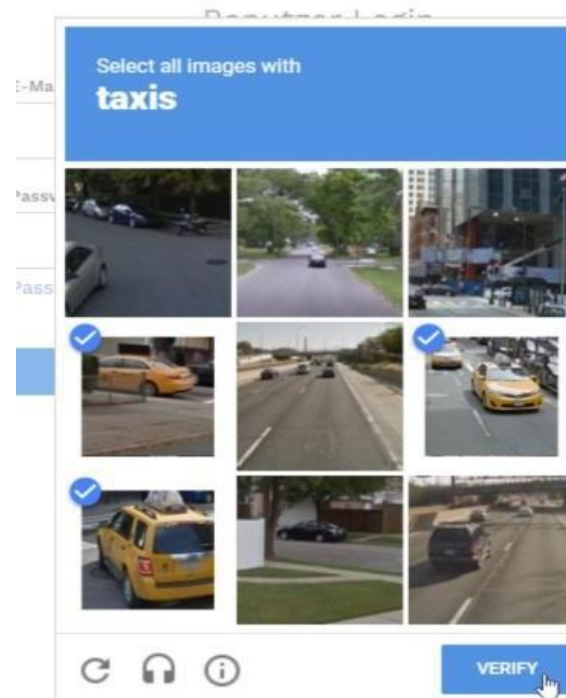


To love and care for others



Selecting all images with traffic lights

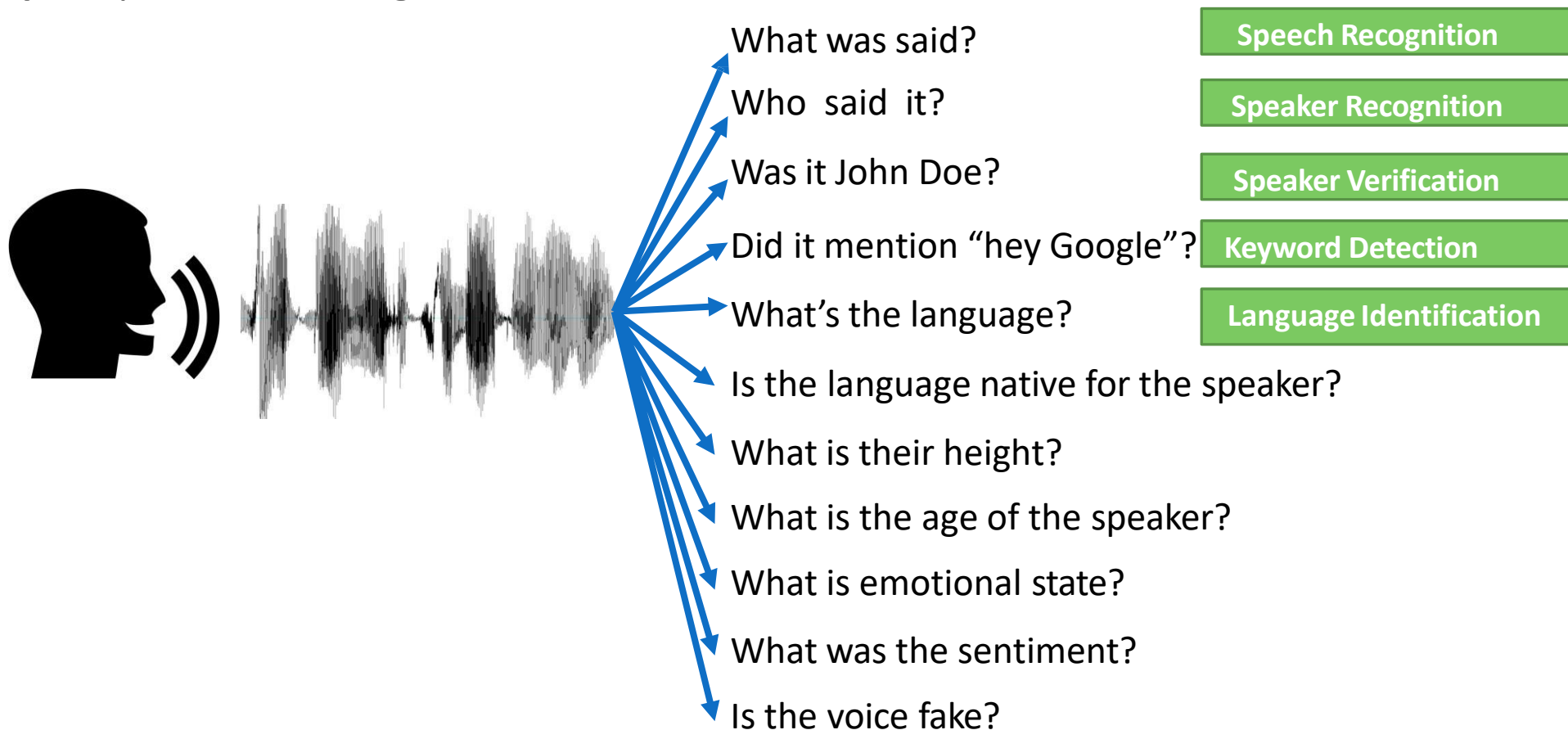
Recaptcha



Task-Label Relationship

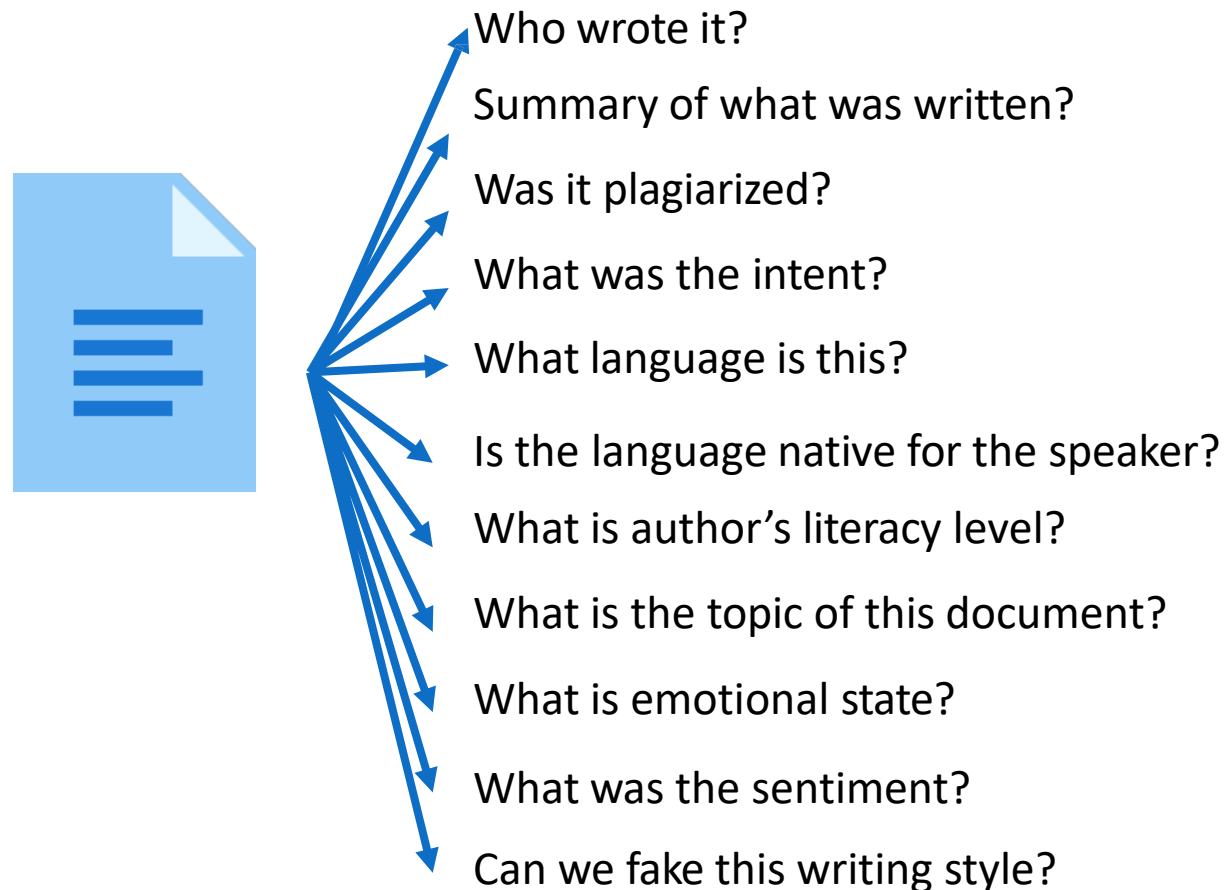
□ Labels are dictated by the task to be performed.

□ **Example:** Speech Technologies



Task-Label Relationship

□ **Example:** Text Technologies



Challenges of ML - Explainability

- ❑ A classifier can potentially learn to classify on the basis of features not desirable for humans
 - All dogs wearing a collar in the training data while no cat is wearing it – ML just learns to separate based on collar
 - All horse images have a copyrights notice – ML just learns to recognize horses based on the copyrights notice

- ❑ **Explainable ML:** The results should be understandable by humans
 - As opposed to a black-box system

Challenges of ML – Fairness

- ❑ AI tends to reflect the biases of the society
 - Human taggers who mark a recording as misinformation based on accent or gender
 - Court decisions in country that make a rich person's acquittal more likely
 - Automated standardized testing in the US could yield unfavorable results for certain demographic groups
 - AI plays a decision role in hiring decisions, with up to 72% of resumes in the US never being viewed by a human (**Automation Bias**)
 - Decision on immigration, bank loans, credit history checks, criminal profiling

ML in Low-resource settings

- Problems where large datasets and tools are not available
- Natural Language Processing and Speech
 - Pakistan has 71 languages
 - We barely have speech recognition capabilities for Urdu!

The Offline Ones

- ❑ 3.6 billion people worldwide are offline
 - That is 46.6% of the world population
 - 13.4% of the develop world, 53% of the developing world, and 80.9% of the Least Developed Countries are offline*
- ❑ Offline Populations
 - Too poor to afford internet-enabled devices
 - Too remote to access the internet
 - Too low-literate to navigate the mostly text-driven internet
- ❑ 285 million visually impaired individuals

*International Telecommunication Union (ITU), <https://itu.foleon.com/itu/measuring-digital-development/offline-population/> (accessed: Feb 2023)

References

- ❑ Murphy Chapter 1
- ❑ Alpaydin Chapter 1
- ❑ TM Chapter 1
- ❑ Lectures of Andrew Ng., Dr. Ali Raza, and “Machine Learning for Intelligent Systems (CS4780/CS5780)”, Kilian Weinberger.
- ❑ This disclaimer should serve as adequate citation.