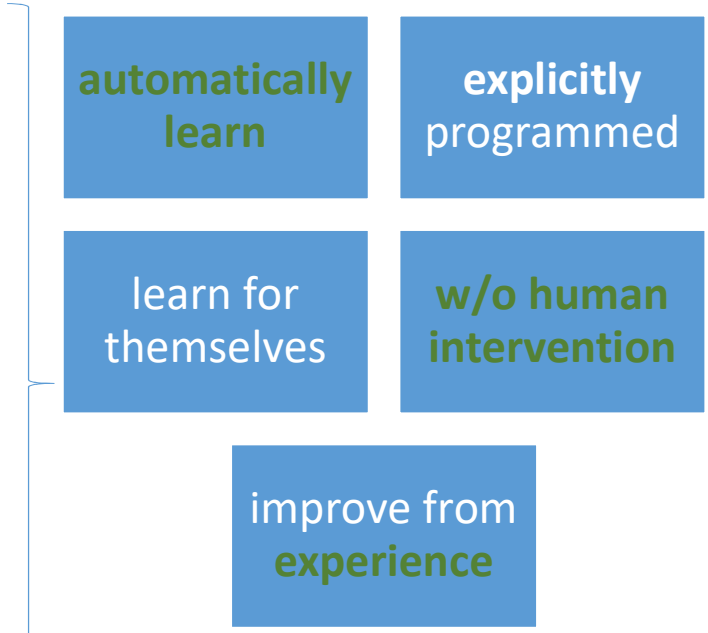
What is ML (still consider both terms are same)

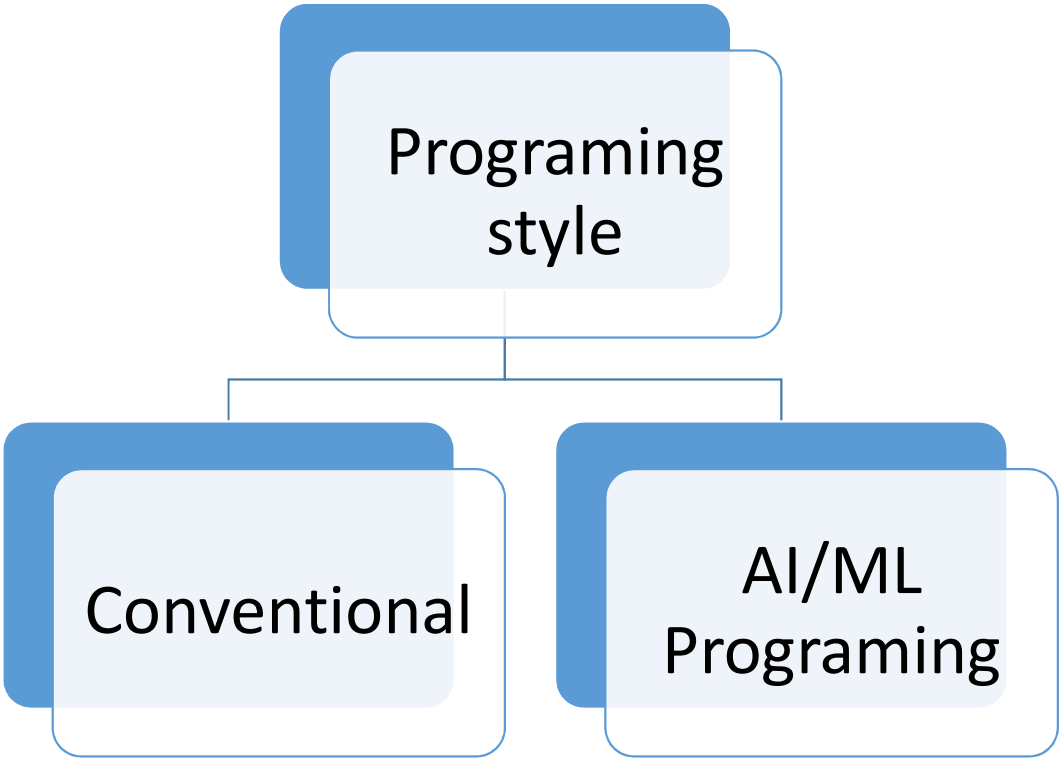
- Machine Learnings provides systems the ability to **automatically learn** and improve from **experience** without being **explicitly programmed**.
- **Machine learning focuses on the development of computer programs** that can access data and use it learn for themselves.
- **The primary aim is to allow the computers learn automatically without human intervention**



Need to understand these terms and its meaning/Interpretation

In other words, this means that the AI learns from experience and makes new decisions independently. It is something that give "*automation of intelligent behavior*" (system change its behaviors intelligently without human interfere, learn by itself, we don't need to program explicitly to adopt the new changes, variations etc of the system).

Understand with example
(What is AIML)



C, C++, JAVA, Python, R, Matlab, OpenCV, …..

Conventional *vs* ML-Programming

Conventional vs ML Programming (Through example)

How to identify if a string is a valid email address or not?

A string of the format **local-part@domain** is a valid email

Conventional way:

- Local-part may use any of these ASCII characters
- uppercase and lowercase Latin letters A to Z and a to z;
- digits 0 to 9;
- special characters not allowed !#$%&'*+-/=?^_`{|}~; (Though this is not the case)
- dot ., provided that it is not the first or last character
- Lets assume white space ‘ ‘ is not allowed
- Domain part

Conventional way:

We have to write a regular expression (regex) based on the above rules to match the string with that pattern. In any high level language this is a really trivial task.

Machine Learning way:

However, when we approach this same task with machine learning, we **do not write** a program to match a ***specific pattern***.

A Machine Learning program is written to find patterns in the **any given dataset** and the **program itself does not know anything** about the type of pattern it needs to identify.

Learn from Dataset- The data needs to be prepared in such a way that it highlights the rules that govern the distinction between what's correct result and what's not.

	Feature Set							Result
	For Local part						Hostname	
Email	ASCII Chars	Only has allov	special chars	dot handling c	space present	at '@' present	hostname valid?	Valid email (Y/N)
hi@codefire.in	Yes	Yes	No	Yes	No	Yes	Yes	Yes
hi.there@codefire.in	Yes	Yes	No	Yes	No	Yes	Yes	Yes
n.o.t.v.e.r.y.c.o.m.m.o.n.9@codefire.in	Yes	Yes	No	Yes	No	Yes	Yes	Yes
user8@localhost	Yes	Yes	No	Yes	No	Yes	Yes	Yes
user@my.solutions	Yes	Yes	No	Yes	No	Yes	Yes	Yes
abc.test2.codefire	Yes	Yes	No	Yes	No	No	No	No
ab c@codefire.in	Yes	No	Yes	Yes	Yes	Yes	Yes	No
hi.there@codefire..in	Yes	Yes	No	Yes	No	Yes	No	No
test~@codefire.in	Yes	No	Yes	Yes	No	Yes	Yes	No

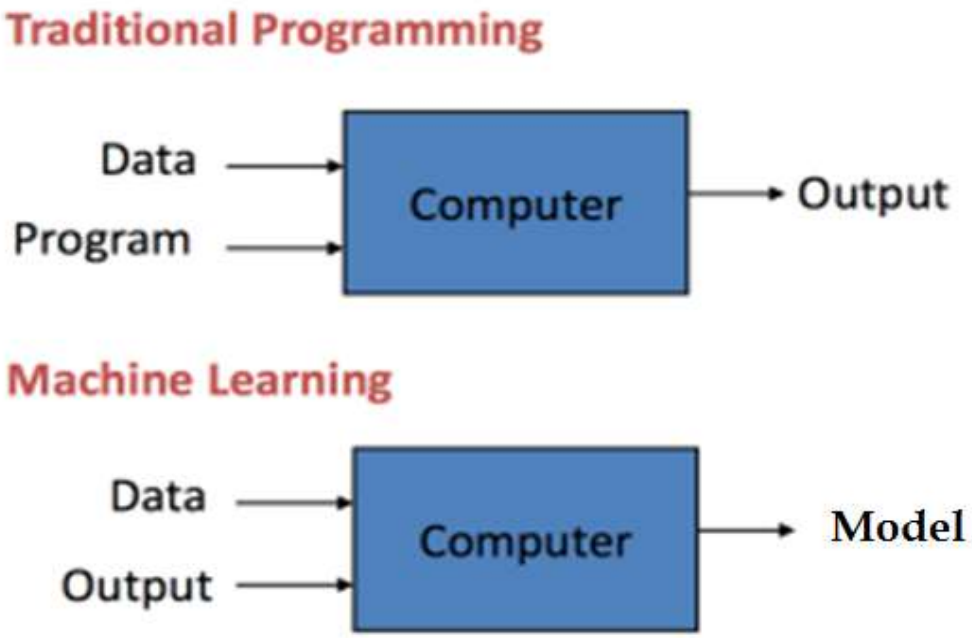
But what happens if the rules to identify email changes?

- Let's say now we allow '~' to be a valid character
- With conventional programming, simply modify the regular expression to allow for ~ to be part of email.
- However, with ML all we will have to do is modify the training data to update the data corresponding to feature set and re-train the ML algo to get the new model.

So in essence:

- 1. ML is a completely different way of looking at a problem.
- 2. ML is applied to problems that do not have clearly defined (fluid) rules. Hence solving that with conventional programing will be more expensive.
- 3. ML programs learn and make decisions based on data, as the data changes the decisions changes. And since humans also learn the same way, by observing what is right and what is wrong around us, hence we term the programs as AI or ML

Need to provide solution **explicitly**



Learn solution **implicitly** (automatically)

Deterministic Vs Non-Deterministic

In **deterministic algorithm**, for a given particular input, the computer will always produce the same output going through the same states (always passing through the same sequence of states), but in case of **non-deterministic algorithm**, for the same input, the computer may produce different output in different runs.

Deterministic	Non-Deterministic
For a particular input the computer will give always same output.	For a particular input the computer will give different output on different execution.
Can solve the problem in polynomial time.	Can't solve the problem in polynomial time.
Can determine the next step of execution.	Cannot determine the next step of execution
No randomness	High degree of randomness
Ex. Calculation of square root, log, addition similar operations	Object recognition

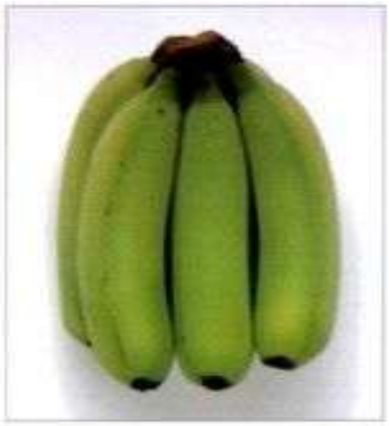
Deterministic Problems

- Program: Addition of two numbers
- Program: Area of a circle
- Program: Maturity amount calculator
- Program: Salary calculation
- Program: Factorial

Non-Deterministic Problems

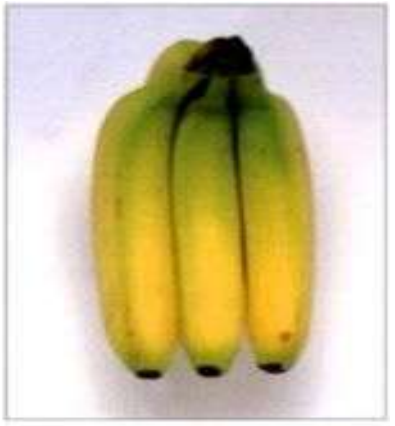
- **Handwritten Digit Recognition**

1. ALL GREEN

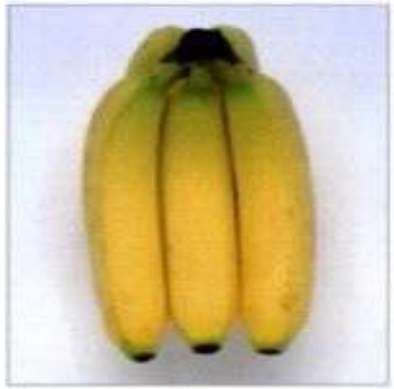


2. GREEN WITH A TRACE OF YELLOW

3. MORE GREEN THAN YELLOW



4. MORE YELLOW THAN GREEN



5. YELLOW WITH A TRACE OF GREEN

6. ALL YELLOW

7. ALL YELLOW WITH BROWN SPECKLES

Pattern Recognition

All Possible (☝) Because of

In 2017, the Economist broadcast that "*Data is the new Oil*"

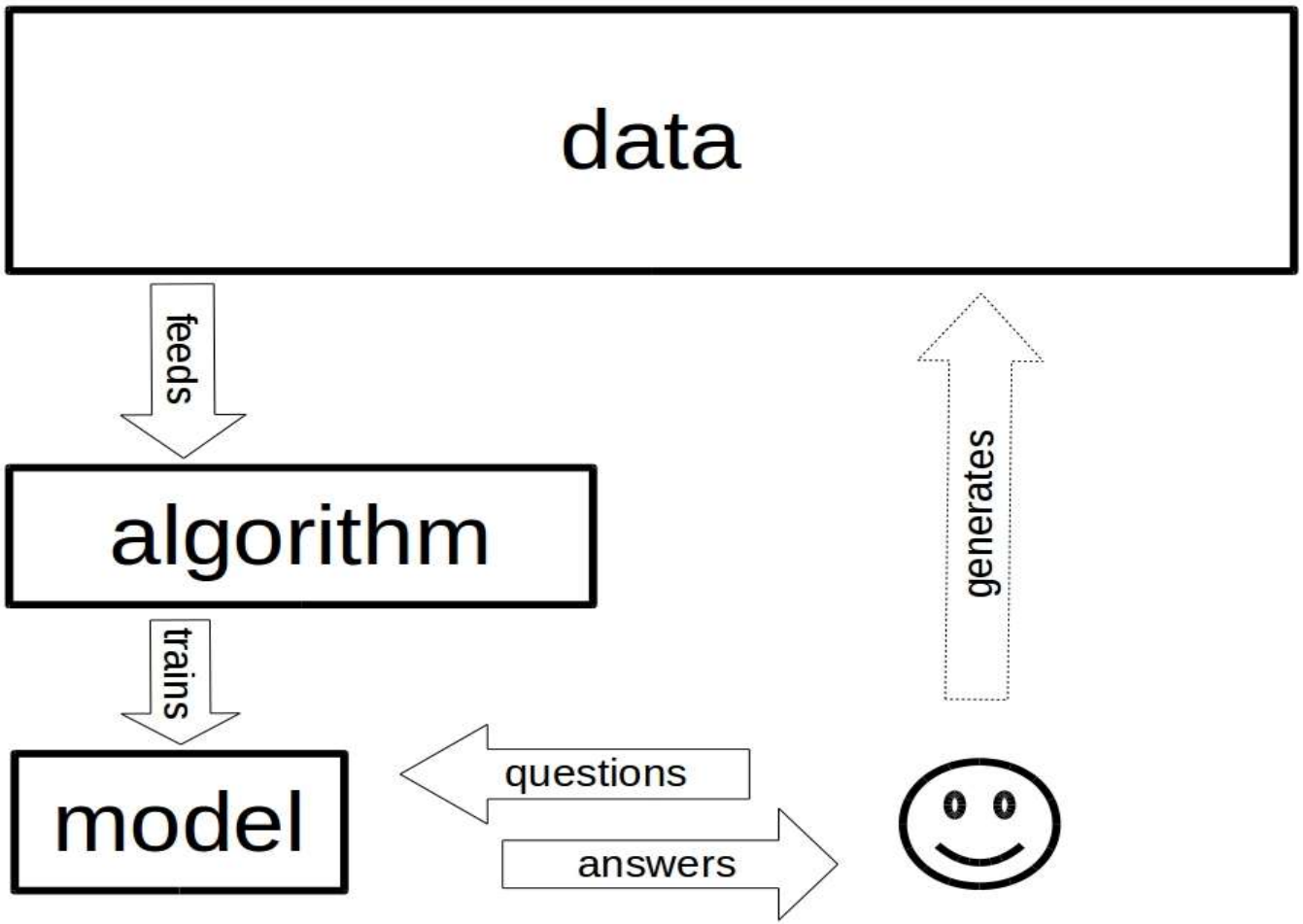
But from the AI and ML point of view, its better to say "*Data is the new Crude Oil*" — in that data must be processed, cleaned and transformed properly before it can deliver real impact.

Machine Learning
Model Development Life Cycle

- Algorithm Selection
- Training
- Model Validation/Evaluation
- Model Optimization

Model

- **Model means trained algorithm which can learn and improve its performance with data.**
- **Model is encapsulation of algorithm and data.**


data
feeds
algorithm
trains
model
questions
answers
generates

Dr. Vivek Tiwari