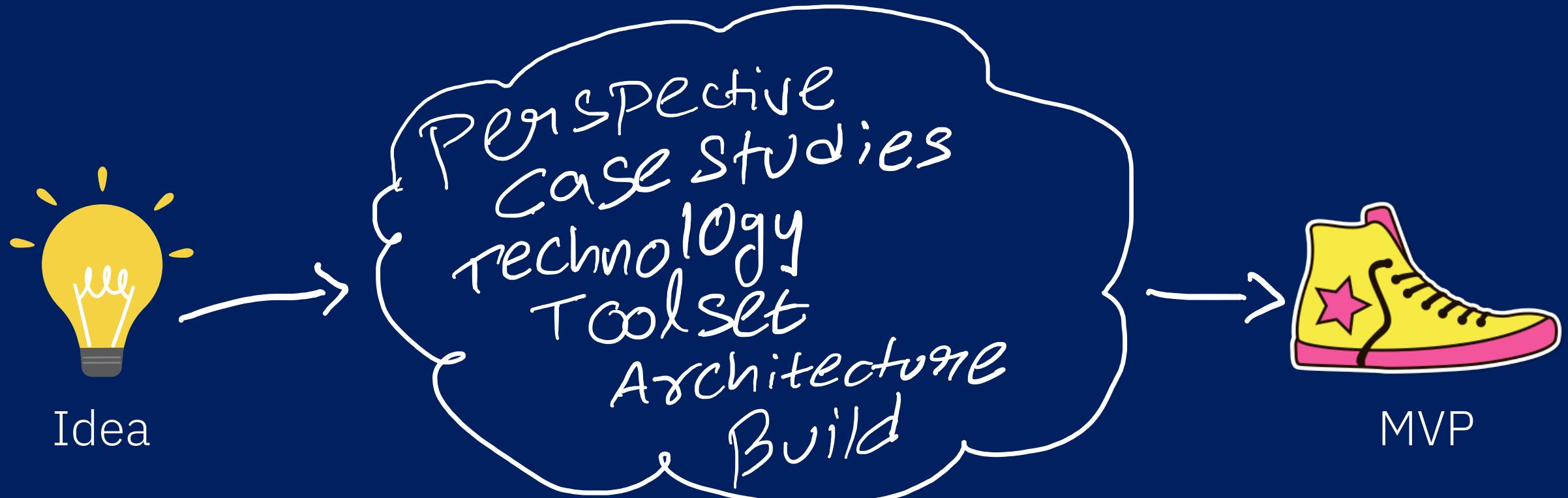


# Enterprise Data science Lifecycle

Krishna Balaga  
DL Engineer/ Developer advocate

# What do i do?



# The Future: Neural nets and Reinforcement Learning

[youtu.be/kHBcVlqpvZ8](https://youtu.be/kHBcVlqpvZ8)

Search for : spot boston dynamics

# Go Fetch!



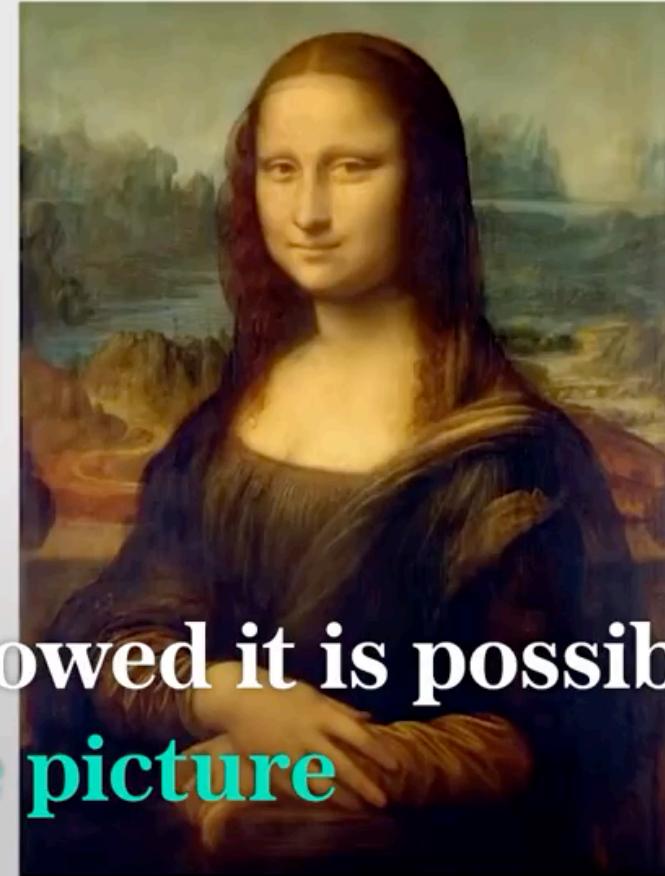
Boston Dynamics

[youtu.be/M8YjvHYbZ9w](https://youtu.be/M8YjvHYbZ9w)

Search for : boston dynamics training

# Bringing Monalisa Back to Life

Living portraits



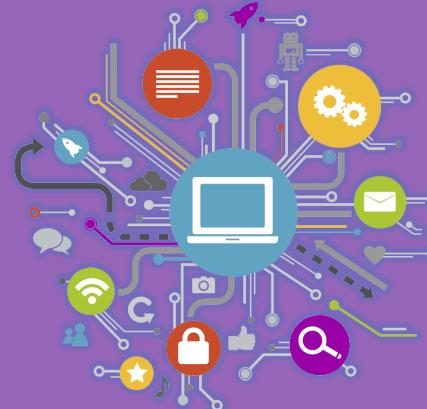
A **Samsung** lab in Russia showed it is possible  
to fake videos from just **one** picture

# Knowledge



Know-how, understanding,  
experience, insight,  
intuition, contextual  
information

# Information



Contextualized, categorized,  
calculated and condensed

# Data



Facts and figures which  
relay something specific,  
but which are not organized  
in any way

## Structured Data

Data with addressable elements

Formats: CSV,XLSX

## Semi Structured Data

Formats:  
XML,HTML

## Un-Structured Data

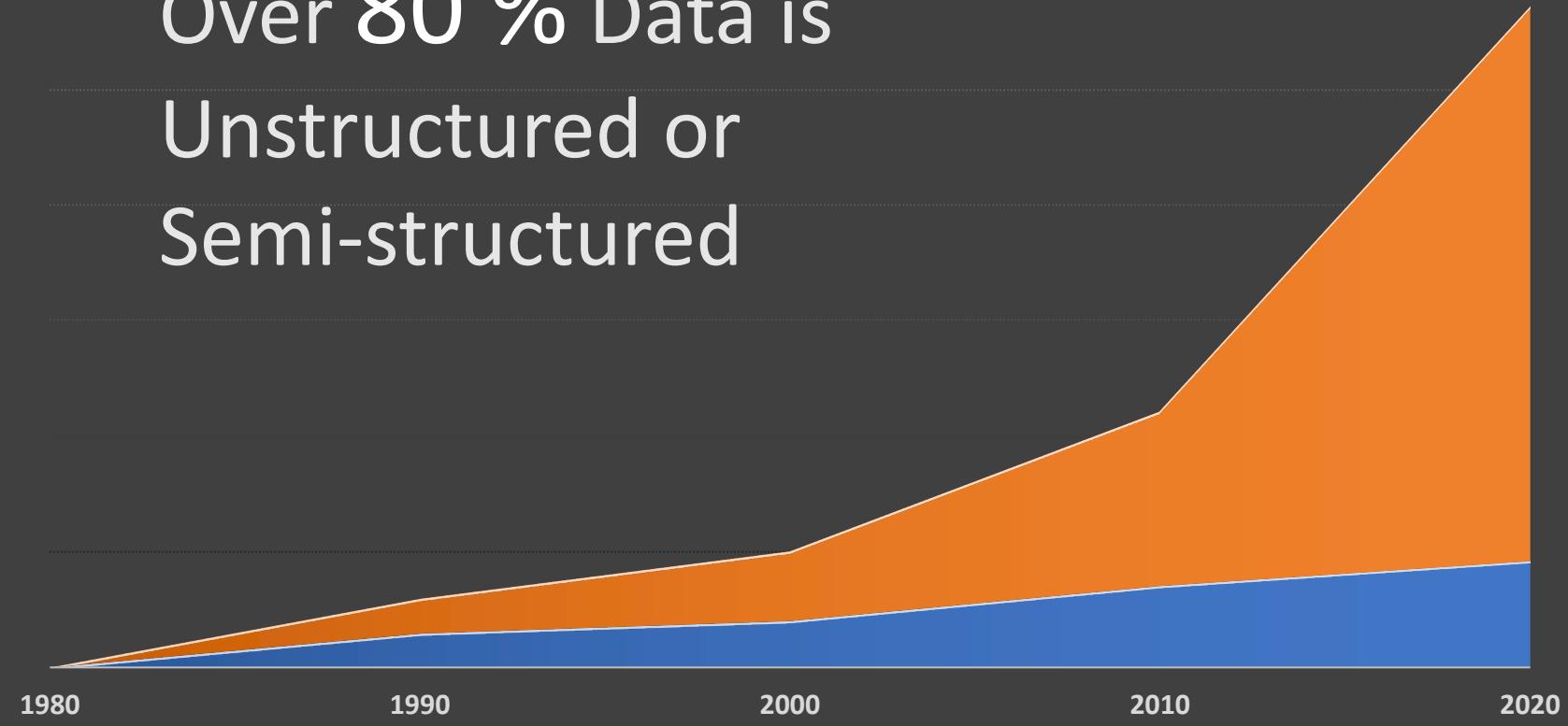
Data without any predefined organization

Formats: JPEG,MP4

## Data Trends over the Decades

■ Structured Data   ■ Un-Structured Data

Over 80 % Data is  
Unstructured or  
Semi-structured





Most BI And Analytics techniques fail in the face of Semi and Un-Structured Data

# Data Science:

A field of study that aims to use a scientific approach to extract meaning and insights from data.

# Machine Learning:

A group of techniques used by data scientists that allow computers to learn from data without programming explicit rules.

# NARROW ARTIFICIAL INTELLIGENCE

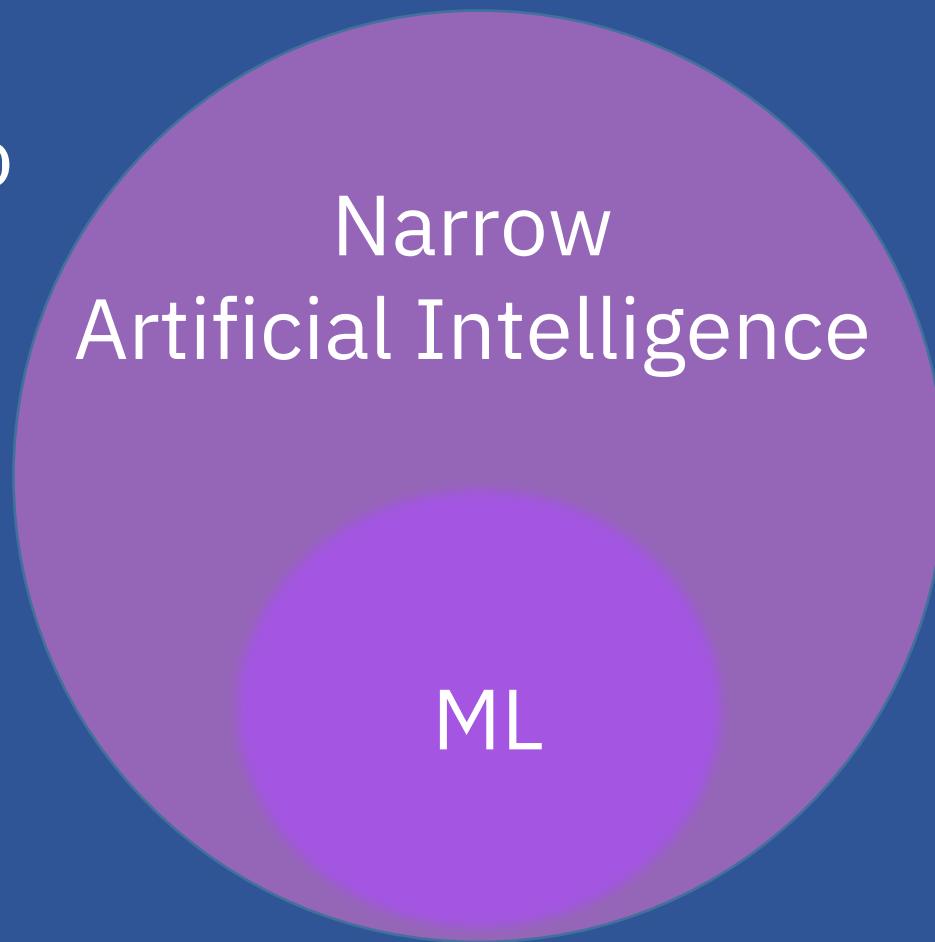
Focuses on a single subset of cognitive abilities  
and advances in that spectrum



Machine  
Learning

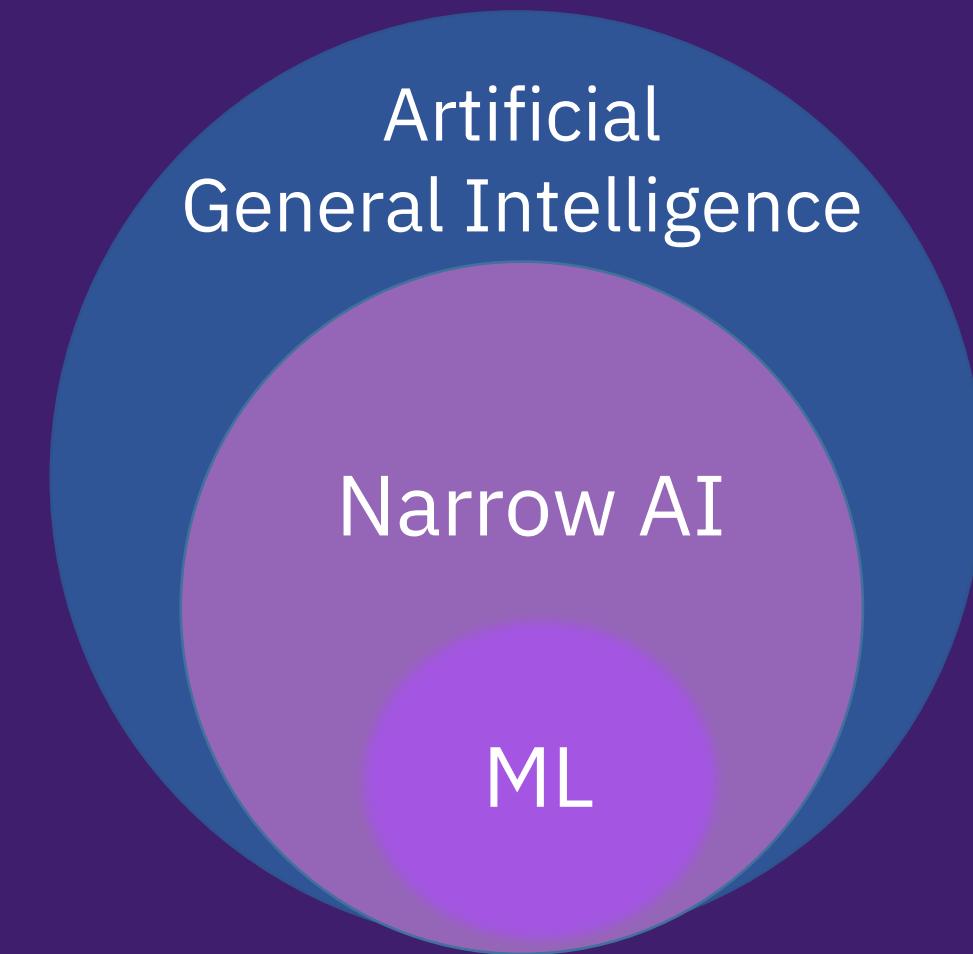
# ARTIFICIAL GENERAL INTELLIGENCE

Intelligence that is as broad as the human cognitive systems, that can-do different kinds of tasks well, and can really simulate the breadth of the human intellect



# SUPER INTELLIGENCE

An intelligence system that rapidly increases its intelligence in a short time, specifically, to surpass the cognitive capability of the average human being



# Super Intelligence

## Artificial General Intelligence

### Narrow Artificial Intelligence

- Median expert prediction For ASI(2060)
- Median expert prediction For AGI(2040)
- Today

# Case Study



Map the atmosphere...  
**every 15 minutes.**

**400 terabytes** of data daily.

More than **275,000 stations** globally.

**100 million smartphone** barometric pressure reports everyday



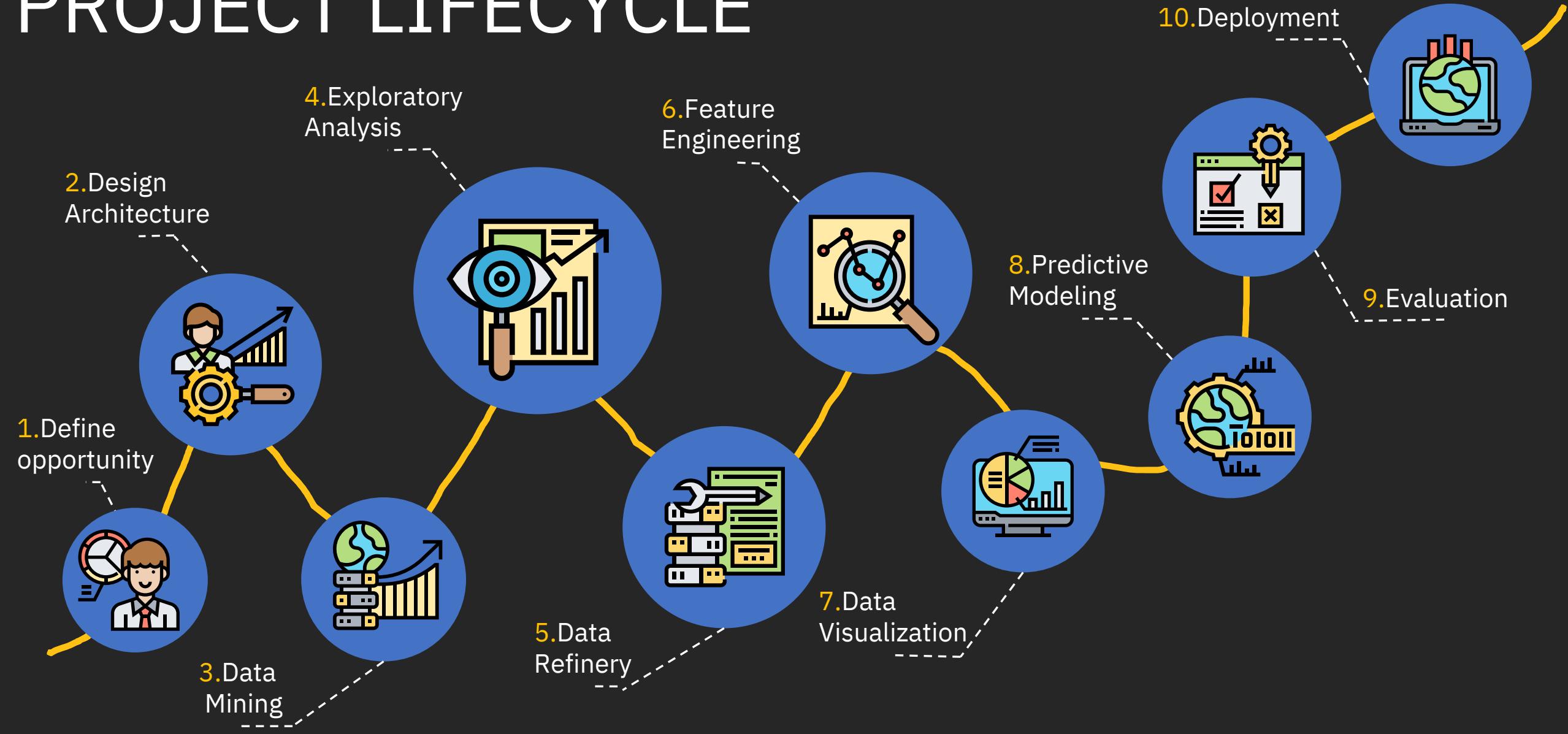
→ **25 billion** forecasts a day.

→ **162** individual forecast models

→ **80%** of the world's top airlines

→ Apple, Facebook, Google,  
Microsoft, Samsung, Twitter

# ENTERPRISE DATA SCIENCE PROJECT LIFECYCLE



# Business Analyst

- Identifies an Opportunity or Problem
- Define the constraints
- Acts as a bridge between Tech and non-tech teams



# Data Architect

- Designs the entire pipeline and defines tech requirements
- Makes decisions on the tech stack being used and compatibility



# DS Manager

- Manages the project
- Forms the required team
- Defines sprints and milestones
- Assigns deliverables



# Data Engineer

- Creates and maintains the required Infrastructure
- Handles data collection , storage and query
- Takes care of Logging and monitoring



# Data Analyst

- Performs basic statistical data analysis
- Does Data profiling and cleaning
- Helps with Data preprocessing
- Generates Basic visualization
- Performs Feature engineering with the help of an SME



# ML Engineer

- Builds the model pipeline
- Performs Feature Transformation
- Takes care of Model selection
- Performs Training and evaluation
- Deployment and scoring



# Data scientist

- Can morph into any of the above Roles involved
- Conducts Research and takes part in  
Creating new algorithms and techniques



# Statistician

- Purely participates in Research
- Helps in solving unconventional challenges



Business  
Analyst

Data  
Architect

Data  
Engineer

Data  
Analyst

ML  
engineer

Define Oppt

Define  
Architecture

Data  
Mining/storing

Exploratory Analysis

Predictive Modeling

Data Refinery

Evaluation

Feature Engineering

Deployment

Data Visualization

# Evolving Roles



# Data Engineer

- Data governance
- Industry wise data compliance

# Data Analyst

- Data Bias detection

# ML Engineer

- Model Bias Detection
- Explainable Models

# Tools of the Future

# Thank you!

Slides: [Github.com/krishnac7](https://github.com/krishnac7)

Linkedin: Krishna Balaga  
[krbalaga@in.ibm.com](mailto:krbalaga@in.ibm.com)