AI and Analytics : Opportunities Landscape

- Srimugunthan Dhandapani
  Data scientist – Technical Lead

Outline

1

AI and Analytics: Introduction

Introduction

2

AI trends

Overview of AI trends and economic opportunities

3

AI/ML applications to different industries

The different kinds of AI/ML problems across industries

4

Data analytics – Roles and careers

Career roles and career opportunities in a data driven company

5

Data scientist Role

Data scientist role and how to prepare a career in data science

.

AI and analytics – Introduction

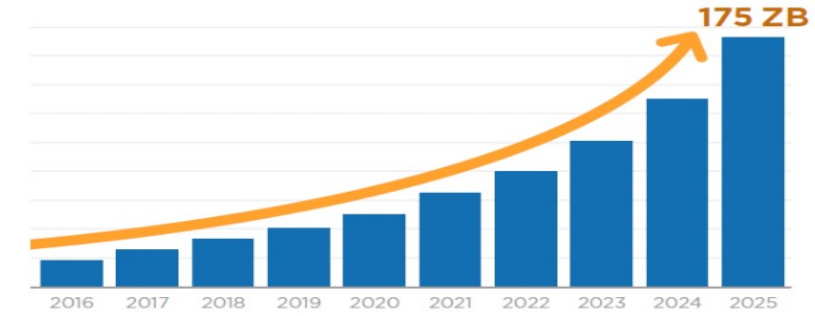
Advent of Big data analytics

DRIVERS OF DATA GROWTH

- Digital transformation of businesses
- Proliferation of devices
- Logarithmic drop of storage costs

EXPLOSION OF DATA

- Data collected mobiles, sensors, and IoT devices
- Data generated in enterprise-hardened servers
- Applications embed data collection as part

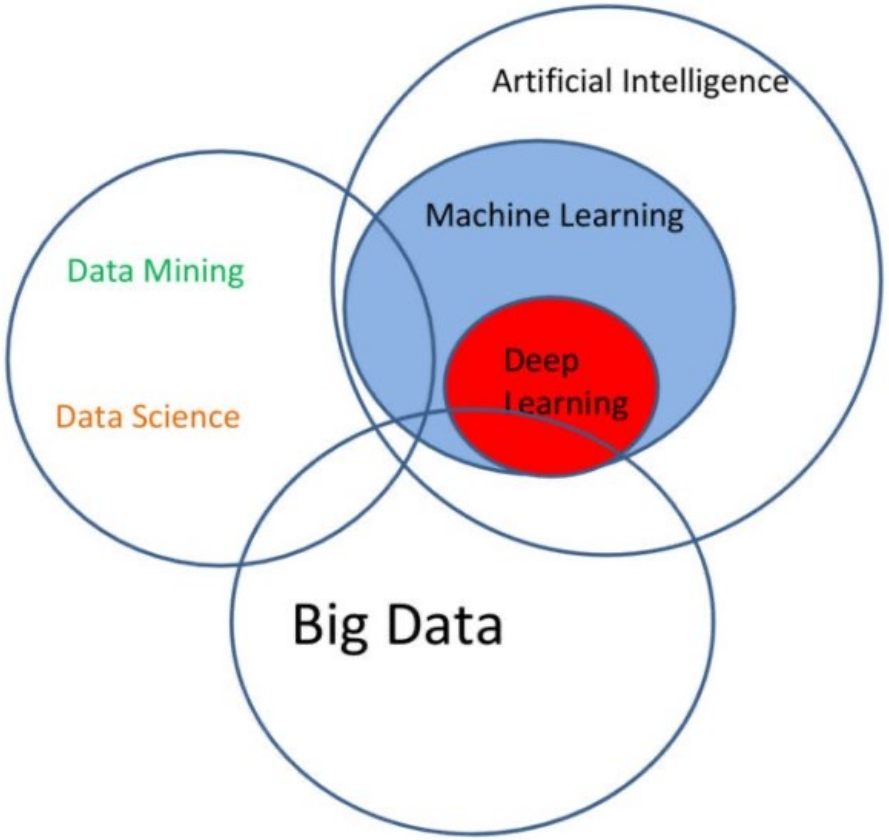


IDC says 175 ZB will be created by 2025 (Image courtesy IDC)

BIG DATA TIMELINE

- 1997: Google launches search engine
- 1999: ”Big data” term used first
- 2005: “Hadoop” paper
- 2010: Spark paper
- 2012 : Role and JD for data scientist became popular
- 2015-2018: Bigdata on the cloud

AI, Machine learning and Big data analytics



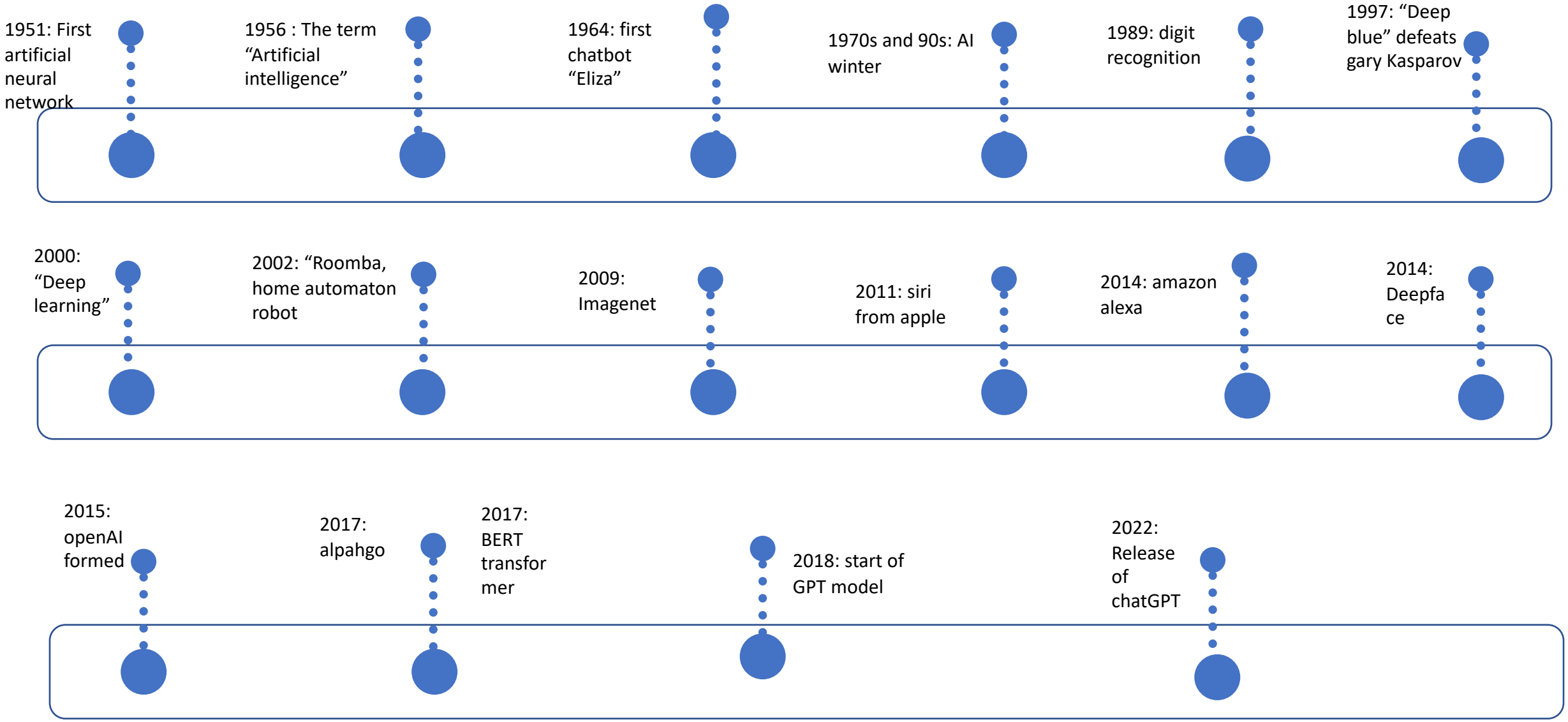
Artificial intelligence: technology that mimics human intelligence

- visual perception
- speech recognition
- natural language understanding.
- automated decision-making

AI techniques include machine learning, deep learning, natural language processing, computer vision, and robotics.

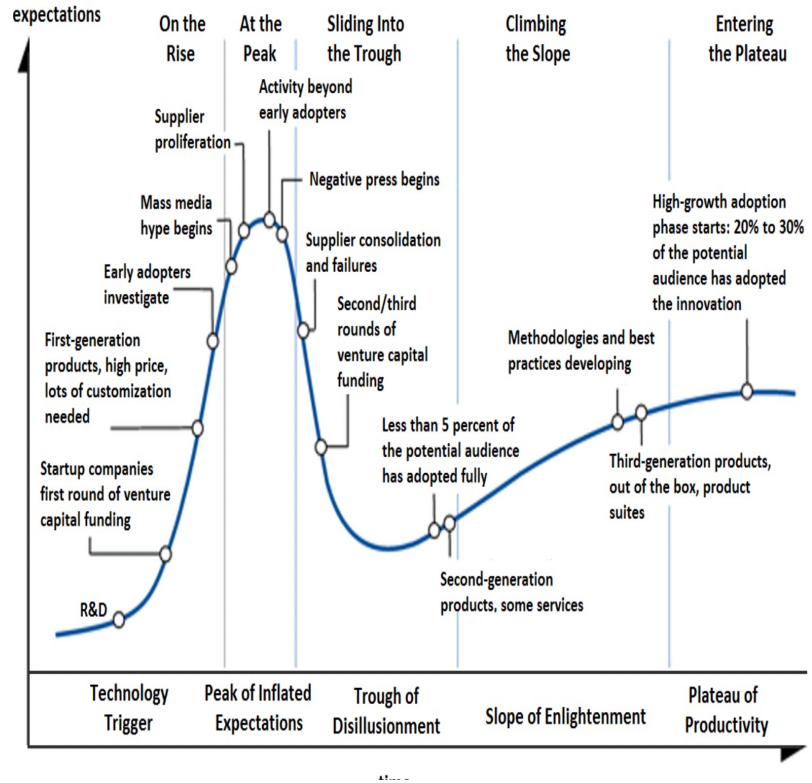
Big data is the fuel on which artificial intelligence runs

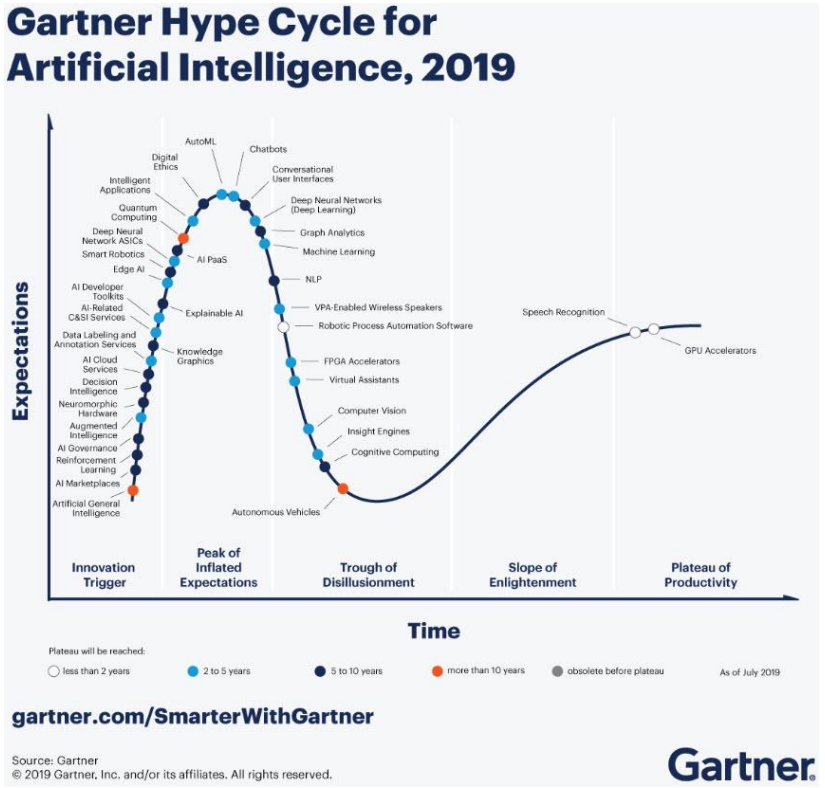
AI machine learning - Timeline

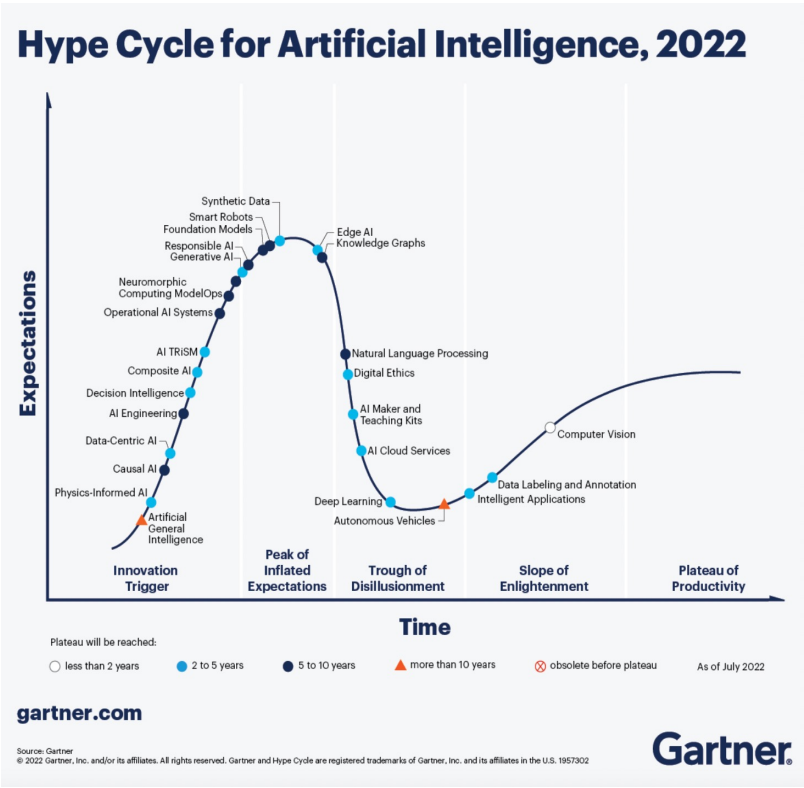


AI trends

AI Trends: Gartner hype cycle for AI







AI Trends

Deep Learning

- relies on the technique of neural networks usually with many layers
- Neural nets can model any function
- With increasing data, the performance keeps increasing, uncovers complex patterns
- Many applications : robotics, Cybersecurity, Smart agriculture

PyTorch

Keras

Computer vision

- Significant progress in Computer vision problems: object detection, tracking, face recognition, image segmentation etc
- Many applications: agri-tech, autonomous cars, contactless retail etc

TorchVision

OpenCV

Chatbots/NLP

- Significant progress in generating next best response
- Improvements needed in terms of contextual awareness, multilingual, nuances in human language
- Applications in customer service, teaching, marketing

Hugging Face

RASA

ChatterBot
Machine learning, conversational dialog engine.

OpenAI

AI Trends

Synthetic data

- lower-quality substitute and used when real data is inconvenient
- banks and financial services institutions use synthetic data by setting up multiagent simulations to explore market behaviours

Data labelling/annotation

- process of attributing, tagging, or **labeling data** to help machine learning algorithms understand and classify the information
- -image, text and audio annotation
- - Semi supervised learning, unsupervised learning

snorkel

Labelbox

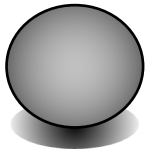
Generative AI and LLMs

- A broad label description of AI that uses unsupervised learning algorithms to create images, video, audio, text or code
- LLM, or Language Model with Latent Variables, is a type of probabilistic model used in Natural Language Processing (NLP) that extends traditional language models by incorporating hidden, or latent, variables into the model.
- many applications in a variety of fields. In the entertainment industry, gaming industry, healthcare industry, marketing etc

Hugging Face

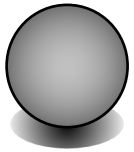
/gpt4-x-alpaca

AI Trends



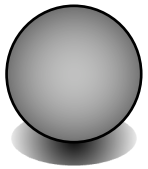
Knowledge graphs

- KG is a kind of semantic network with added constraints
- encode human knowledge leveraging a graph-based structure
- machine learning models will have better explainability and trustworthiness.
- Applications: Finance domain for fraud detection, social media, new product analytics



Smart robots

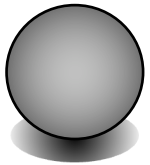
- learning about their environment and experience.
- better customer experience for doing repetitive tasks
- Applications: Home assistant robots,
- Service robots in hospitality, restaurants etc



Causal Machine learning

- AI is good at predicting “when X then Y”. However causal ML tries to understand why
- Some of the key challenges in causal machine learning include dealing with selection bias, identifying appropriate causal assumptions, and handling complex causal structures.
- Causal machine learning algorithms typically use observational or experimental data to infer causal relationships between variables

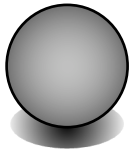
AI Trends



AI trism

- AI trust, risk and security management (AI TRiSM) ensures AI model governance, trustworthiness, fairness, reliability, robustness, efficacy and data protection. This includes solutions and techniques for model interpretability and explainability
- Manage AI risk are much more likely to experience negative AI outcomes and breaches. Prevent financial and reputational loss

- Avoid Models not performing as intended
- Compliance to AI regulations
- Manage security and privacy failures



AI governance

1.Lifecycle management of AI models
2.MLOPS challenge: sensing and interpret a poorly working AI system. Designing alerts for problems
3. continuously learning AI models and monitor learning

Other ML topics

- Hardware accelerators:
- Neuromorphic computing
- Reinforcement learning
- Self supervised learning
- Physics informed neural nets

AI Tools

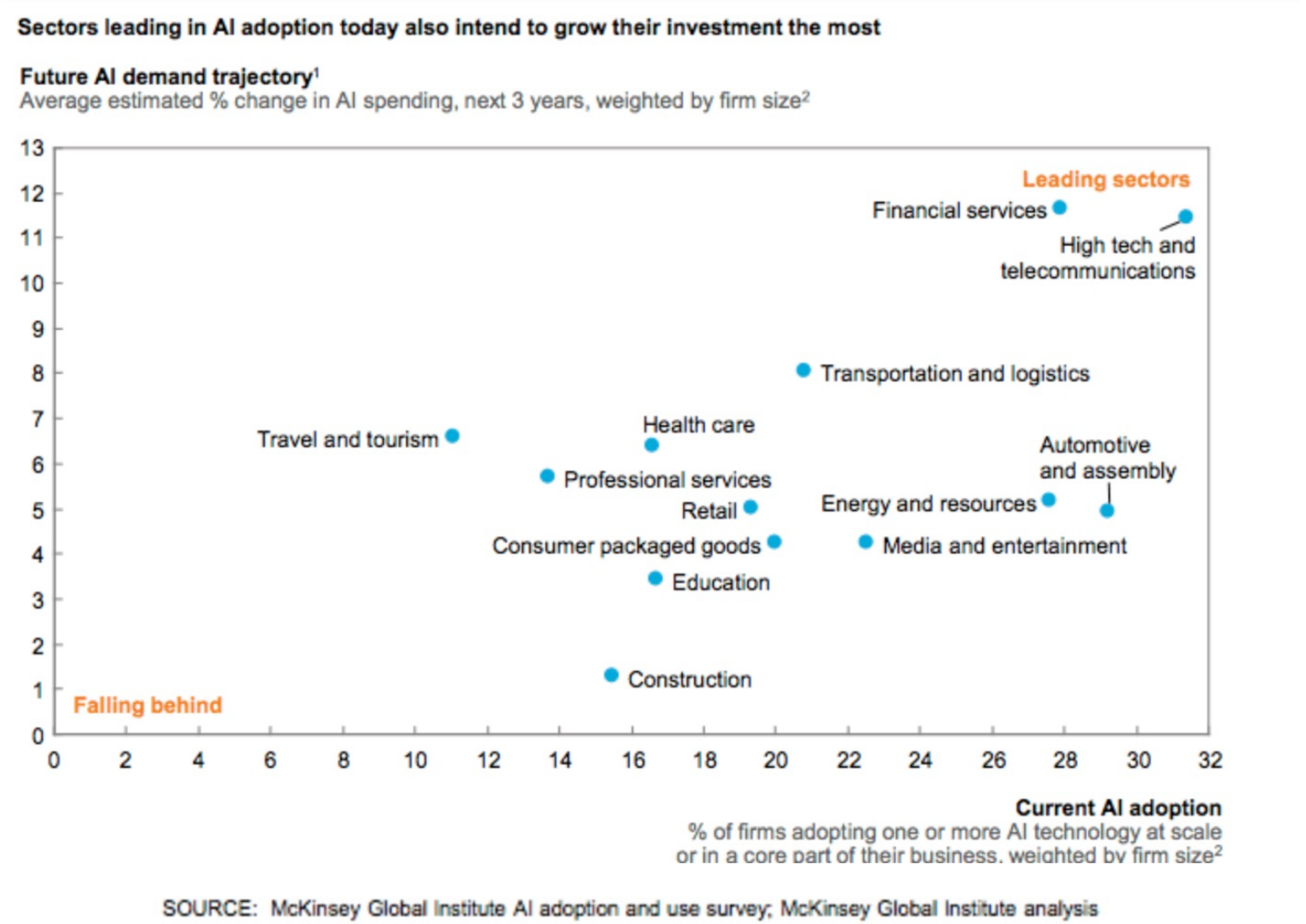
Writing assistants	Jasper
Meeting assistants	fireflies.ai, Otter.ai
Learning & Research tools	Teach Anything, www.explainpaper.com Explainpaper
Note taking tool	Project Tailwind, Notion AI
Marketing	SURFER, toonly, FlexClip
Customer support	chatfuel
Schedule and plan	JourneyPlan
Presentation	tome, beautiful.ai
Translation: Text-to-speech Text-to-music Text-to-image	MURF.AI, SOUNDRAW, DALLE-2

A comprehensive collection of AI tools : https://www.futuretools.io/

AI/ML Applications in different industries

AI in different sectors

- All economic sectors adopt AI
- Driving value by
  - Decreasing expenses
  - Innovation: Help Launch new products or services
  - Improved decision making
  - Improved productivity
  - Enhancing Reliability security



AI/ML problems in different sectors

	SECTOR	AI/ML APPLICATION
	Telecom	• How to detect churn ? • How to manage the network infrastructure effectively and proactively find problems? • How to segment customers based on value? How to predict customer lifetime value? • Dynamic pricing, promotion • How to optimize call center operations?
	Financial services	• Risk modelling • Fraud detection • Algorithmic trading • Robo-advisors
	Retail	• Inventory and merchandising management • Personalization

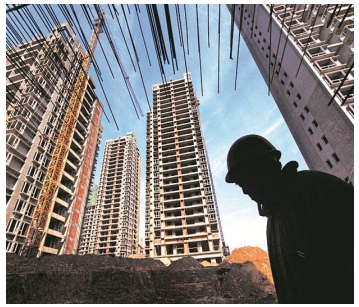
AI/ML problems in different sectors

	SECTOR	AI/ML APPLICATION
	Manufacturing	• Predictive maintenance • Robotic process automaton • Industrial IOT applications
	Agriculture	Precision farming Crop monitoring Soil analysis and monitoring Dynamic pricing
	Energy	Demand prediction and optimize energy grids Decisions for where to drill for oil from seismic data

AI/ML problems in different sectors

	SECTOR	AI/ML APPLICATION
	Travel tech & Transportation	• Travel assistant • UX personalization • Flight fare/hotel pricing • Matching riders to drivers • Travel route optimization • Dynamic pricing
	Health care	• Disease diagnosis from medical reports and images • Drug discovery • Gene analysis and editing
	Education	• Customized learning journeys • Personalization • Content creation

AI/ML problems in different sectors

	SECTOR	AI/ML APPLICATION
	Construction/Real estate	• Building information modelling • Predict housing prices
	Consumer Tech	• Recommendation for cross selling • Customer purchase funnel optimization
	Media and entertainment	• Content generation and personalized content • AR/VR games

Indian AI startups across domains

AI PLATFORM opslyft

FUTURE CARS FLUX AUTO

DRONES Garuda aerospace

INTELLIGENT ROBOTS Invento ROBOTICS

MEDICAL SIGTUPLE

HEALTH AND WELFARE wysa

SMART MANUFACTURING Flutura Actions! Not just Insights

SMART CITY wobot Intelligence

AGRICULTURE Cropin

FINTECH LENDINGKART Simplifying MSME Finance

EDUTECH EMBIBE

REAL ESTATE NOBROKER

LAW/PATENT SPOTDRAFT

TRAVEL ixigo

RETAIL/ E-COMMERCE InfiniteAnalytics

SECURITY DeepSight AI Labs

INSURANCE policybazaar.com Compare. Buy. Save.

FOOD/RESTATURANT gulpie

MEDIA/ENTERTAINMENT pratilipi

DEVICES/IOT Uncanny Vision®

CANTACT CENTER SquadStack

MARKETING/SALES comzense Bridging Technology & Business

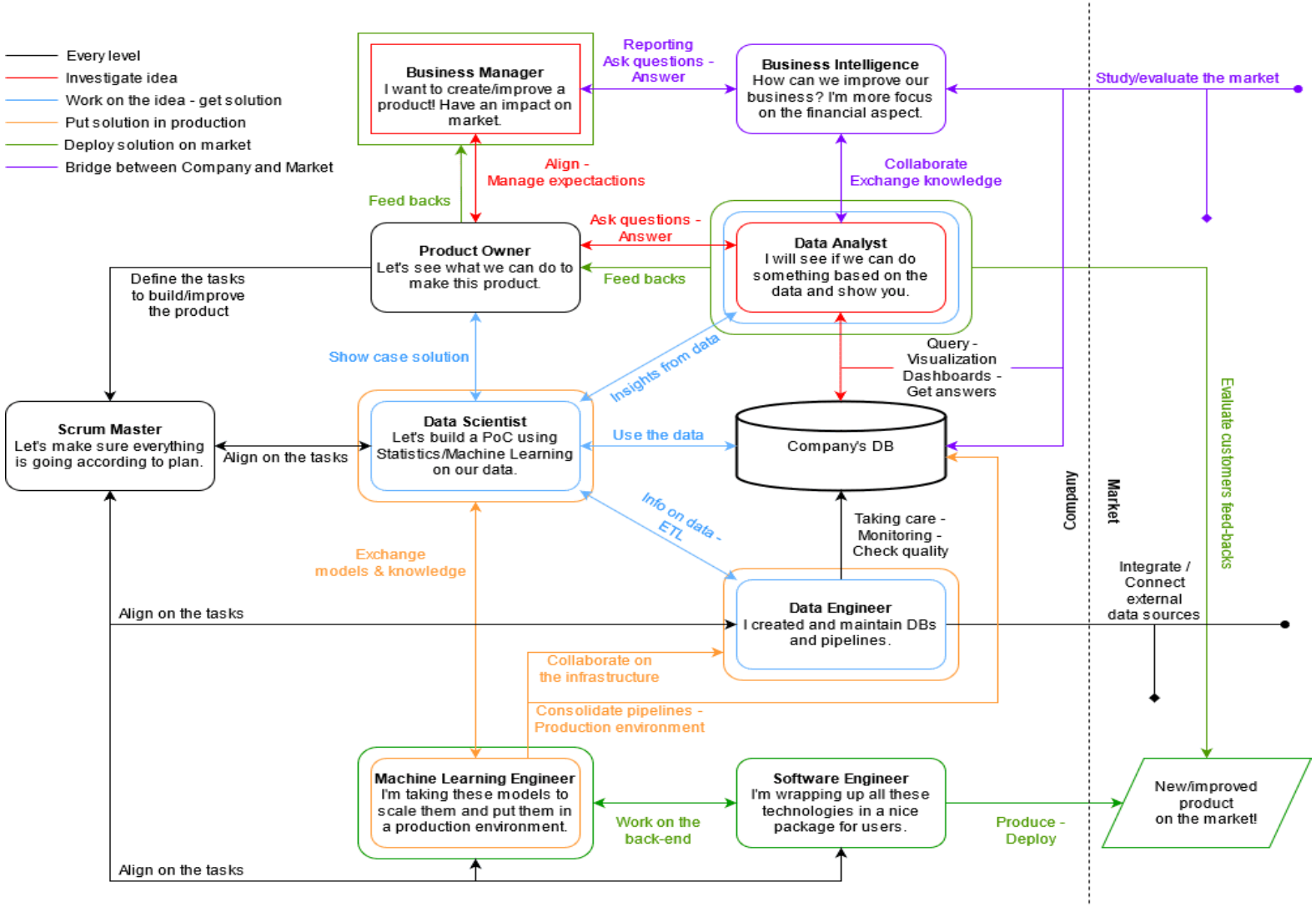
TELECOM Wiom

GAMING NORAH AI

FASHION BIGTHINX

Data driven companies and Data careers and roles

Different roles in creating a data product

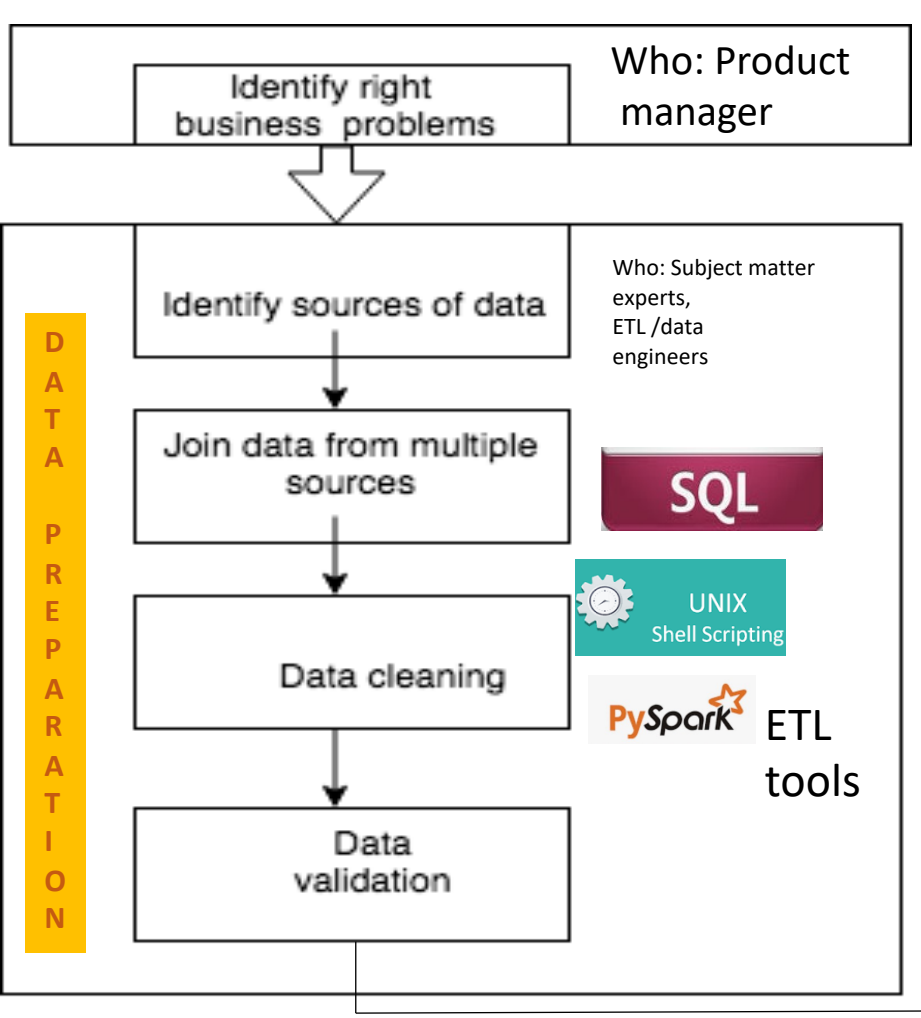


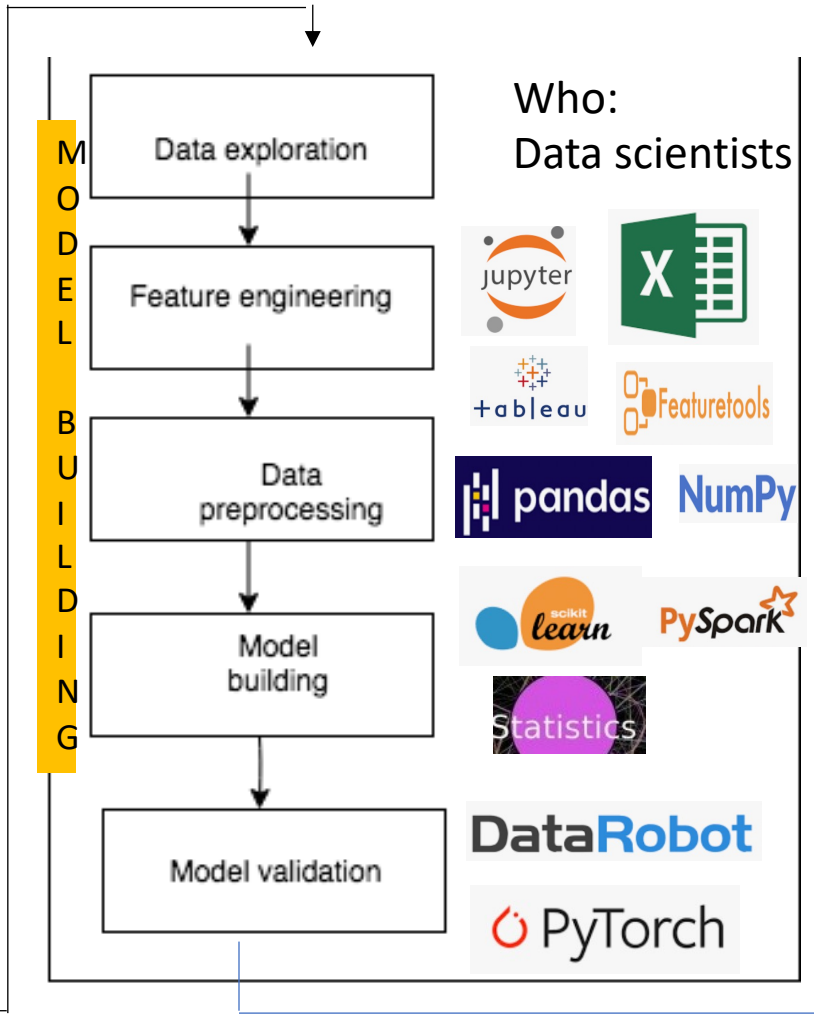
ROLES:

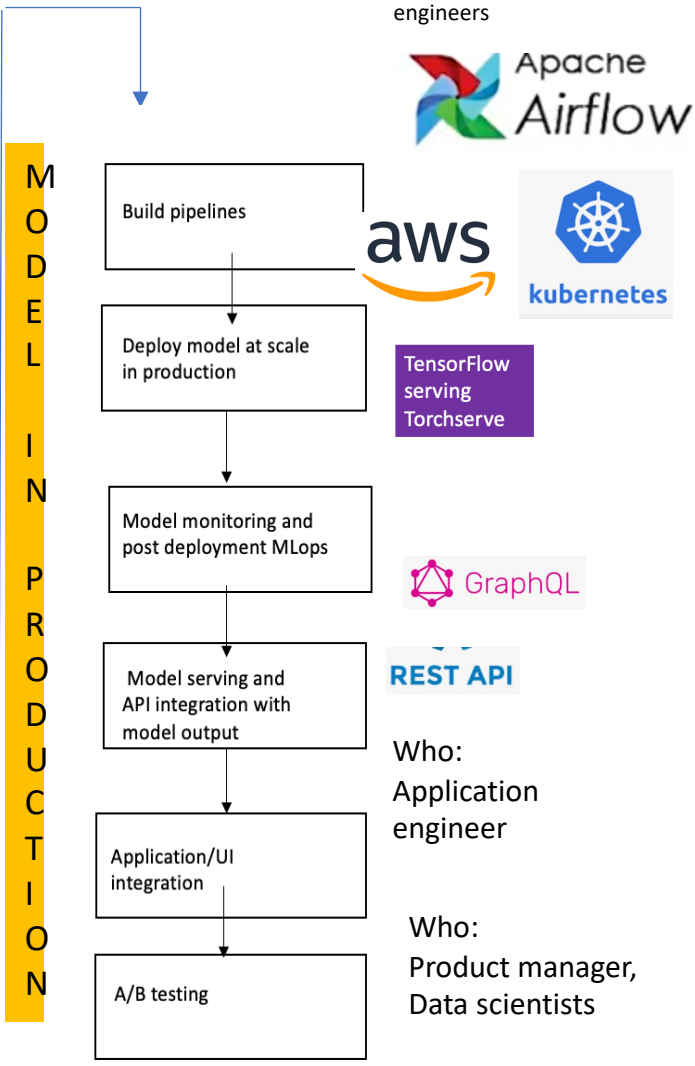
- Business manager
- Business Intelligence
- Product owner
- Scrum master
- Data scientist
- Data analyst
- Data engineer
- ML engineer
- Software engineer

https://www.kdnuggets.com/2021/12/build-solid-data-team.html

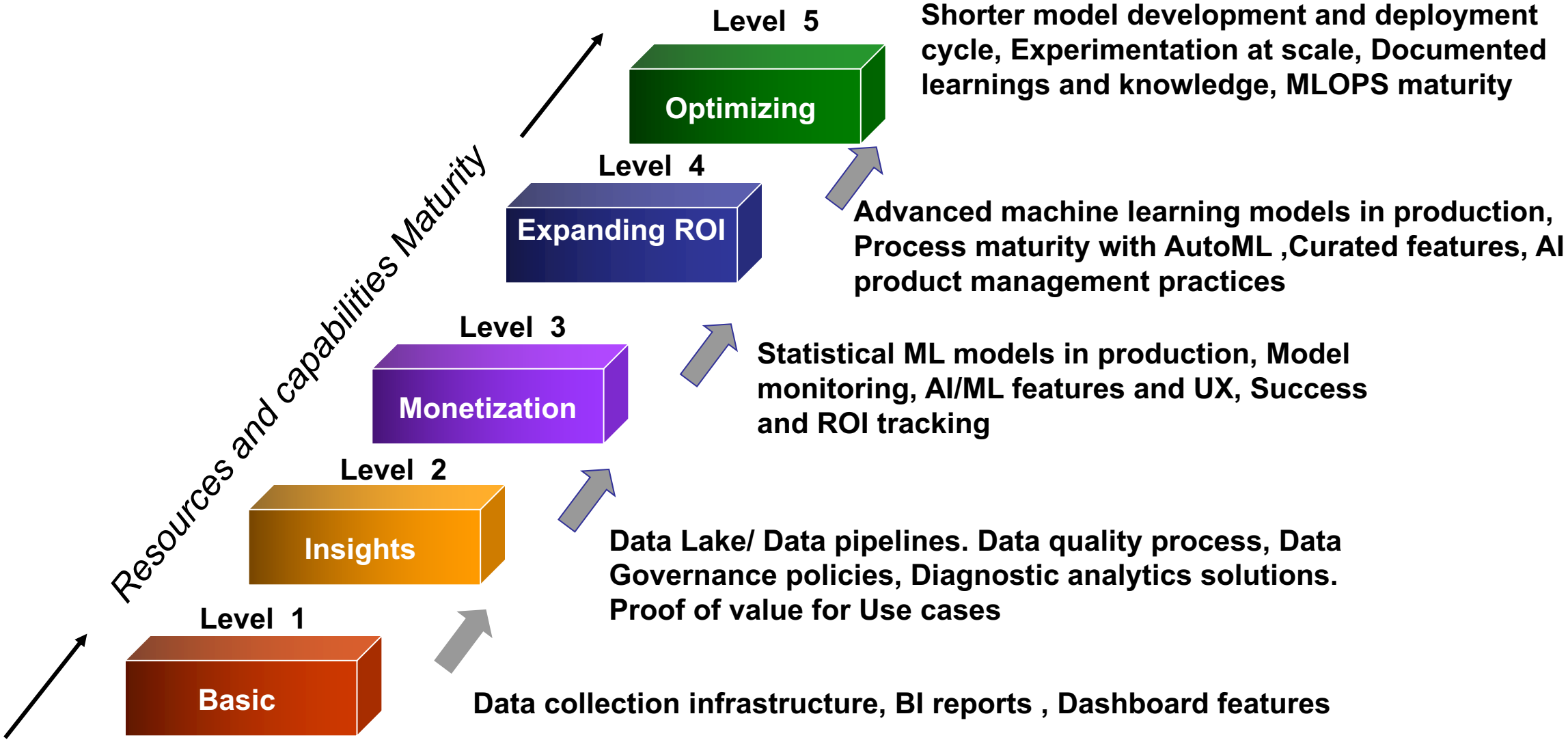
ML model lifecycle







AI/ML org maturity levels



Data and analytics maturity index

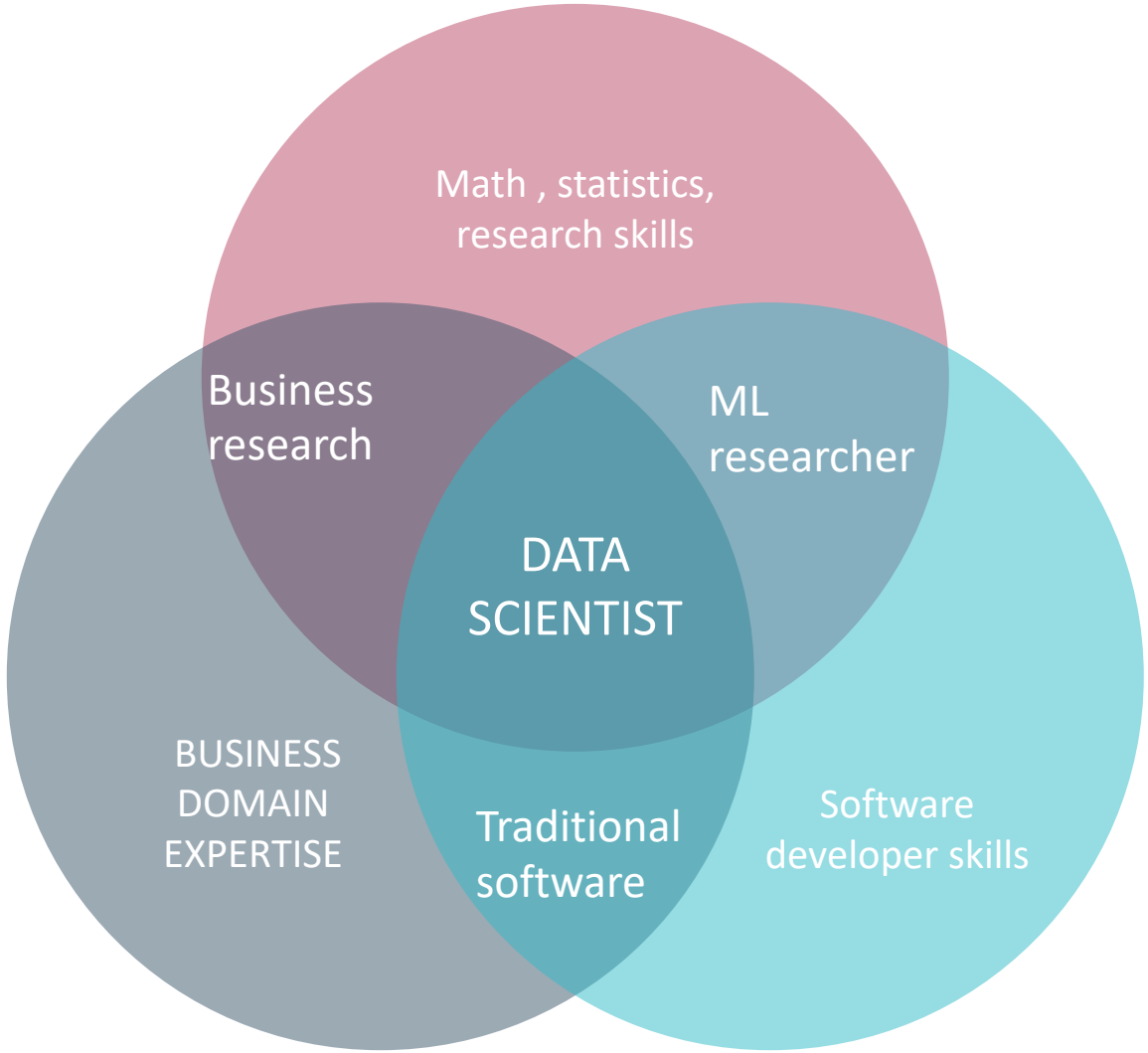
Source: Bill Schmarzo "Big Data MBA" Course Curriculum

Questions that assess maturity:

1.how much data is generated and stored everyday?
2.Where do data reside? What kind of data platform? How many pipelines run everyday?
3. what is the model deployment platform? Do you use in-house infra/cloud ?
4.is there a feature store for curated features
5.how is data quality assessed within the org?
6. how is the data governance policies? how is access controlled to these data
7.what tools you use to discover data?
8.are there tools to find out lineage of a data source or table?
9. what are the tools for model development? what experiment tracking tools do data scientists use?
10.what processes are there to assess model bias?
11.how many A/B test/experiments are done an year?
12.what kind of post model MLops is done? how do you identify drift?
13.do you track the cost of model deployed? how are models retired?
14.how many real time models are deployed per year?
15.what is the ratio for data analysts: data scientists: data engineers: ML engineers?
16.how are the results of data analysis documented and propagated?
17.what is the process and reviews do models go through before deployment?
18.how are the model outputs integrated with UX?
19.how many deep learning models run in production? what are the use cases ?
20. Do you use any AutoML tools for feature engineering and modelling?

Data scientist role

Data scientist role



THREE MAIN EXPERTISE FOR DATA SCIENTIST

1. Software developer skills:
2. Business domain expertise:
3. Math, statistics and research skills

LinkedIn titles with data science role:

- Experimental scientist
- Decision scientist
- NLP engineer
- Computer vision engineer
- AI researcher ..etc.

Data science Knowledge areas

ML Essentials for beginners:

- Linear regression
- Logistic regression
- KNN, LDA
- Ridge and Lasso regression
- Feature selection
- Cross validation and bootstrap
- Bias variance tradeoff
- Decision trees
- Bagging , Random Forest
- Boosting
- SVMs
- PCA
- K-means clustering

and Hierarchical clustering

LEVEL3	Causal inference and causal ML	Simulation	Tuning deep learning and other architecturesRNN, LSTM, CNN	Operations research	Reinforcement learning	LLM	Bayesian approaches	Experiment design
LEVEL2	Anomaly detection	Deep learning	Computer vision	recommender systems	Ranking models	Time series models	NLP	
LEVEL1	ML model deployment Model explainers	Model validation - Bias variance tradeoff - Learning curve Cross validation	Hyper parameter tuning	A/B testing	Feature selection methods	Feature engineering - data transformation - outlier detection - binning		
LEVEL0		SQL	Descriptive statistics	Statistical hypothesis test	Data visualisation	Basic ML algorithms from scratch		

Data scientist resume

Resume header

SKILLS

Project portfolio

Public profiles

FIRST LAST

Bay Area, California · +1-234-456-789 · professionalemail@resumeworded.com · linkedin.com/in/username

Skills

- Python, Pandas, scikit-learn
- Apache-spark
- R
- SQL
- Tableau

Align skills to job description

Projects

Movie Recommendation System

- Used k-nearest-neighbors to build a recommendation model for movies
- The model outperformed achieving 18% improvement over baseline

Attrition model

- Built a employee attrition model using random forest and achieved results with estimated savings of 10 million$

Sales forecast model

- Implemented a time series model to forecast the sales and improved forecast accuracy by 40%

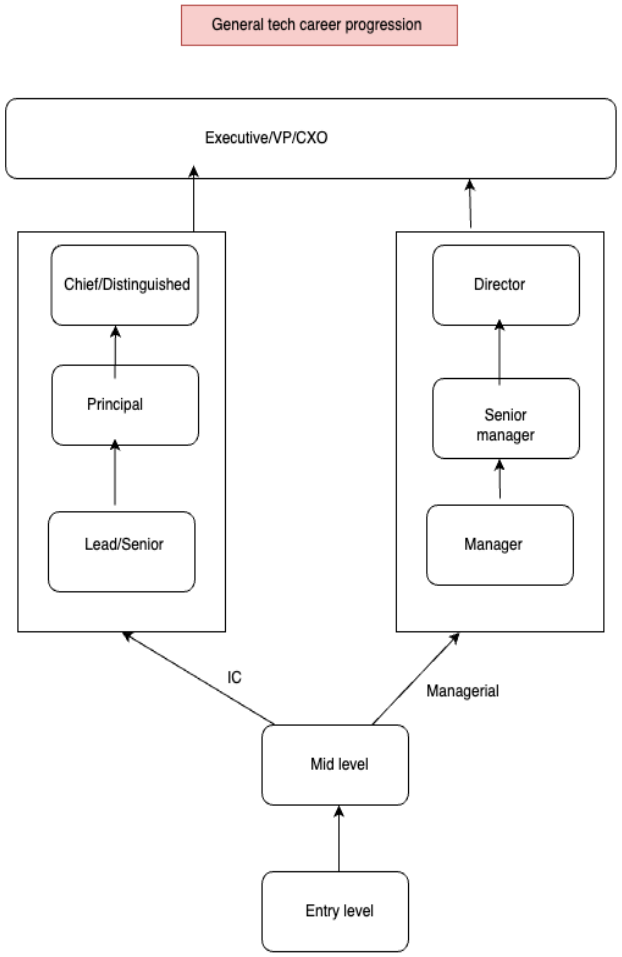
Project portfolio expected to have a regression, classification, recommendation problems along with a data analysis project

Resume bullet point template

Github: https://github.com/xxyhz
Linkedin: https://linkedin.com/in/xxyhz
Blog: https://medium.com/@xxyhz

- GitHub home page
- Technical blog
- Personal website etc

Data scientist – Career progression



IC levels

Walmar...	Amazon	Accenture
Data Scientist 2	DS I L4	Associate Data Scientist
Data Scientist 3 x3	DS II L5	Data Scientist Analyst
Senior Data Scientist x4	DS III L6	Senior Data Scientist
Staff Data Scientist x5		Team Leader
Principal Data Scientist x6	Principal DS L7	
Distinguished Data Scientist x7	Senior Principal DS L8	

https://www.levels.fyi/t/data-scientist?countryId=254

managerial levels

- VP -Data science
- Director -Data science
- Senior DS Manager
- Data science Manager

Thanks

All the best

LinkedIn connect

https://www.linkedin.com/in/srimugunthan-dhandapani/