IBM DS0101EN

What is Data Science?

‘’Every positive jump for humanity has been fuelled by intelligence.“

What is intelligence?

The ability for solving problems

What is artificial intelligence (AI) ?

Artificial Intelligence is a set of computer science techniques that allows computer software

to learn from experience,

adapt to new inputs

And complete tasks that resemble human intelligence.

Why talk AI now?

Computing power is AI's engine
Data is AI's fuel
Algorithms are AI's design

Applications of AI in Various Industries

INDiGROWTH
insights driven growth

Machine learning & AI being used everyday in more ways than ever

ON YOUR SMARTPHONE...

Ok Google

What channel does GoT Air On?

Hey Siri

Hey Cortana

Translate .

Que voulez-vous dire...

Maps

Way from the airport to home

WHEN YOU'RE...

FB Moments

Pics of you & I at Anna's party

Shopping

amazon.com

Customers who bought This utem also ..

Videos

Other movies you might...

Music

pandora

Recommened

Email

Primary inbox, smart reply

MAKING BUSINESS HAPPEN...

Robo-advisor

Betterment

Your Investment Portfolio

Scoring Engine

ETS

Writing Proficiency

Marketing & Advertising

Albert frank.ai

Brining it all together in Real-Time

Fraud Detection

PayPal

Machine Learning at play

Machine Learning Use Cases in Finance

Financial Monitoring

Making Investment Predictions

Process Automation

Secure Transactions

Risk Management

Algorithmic Trading

Financial Advisory

Customer Data Management

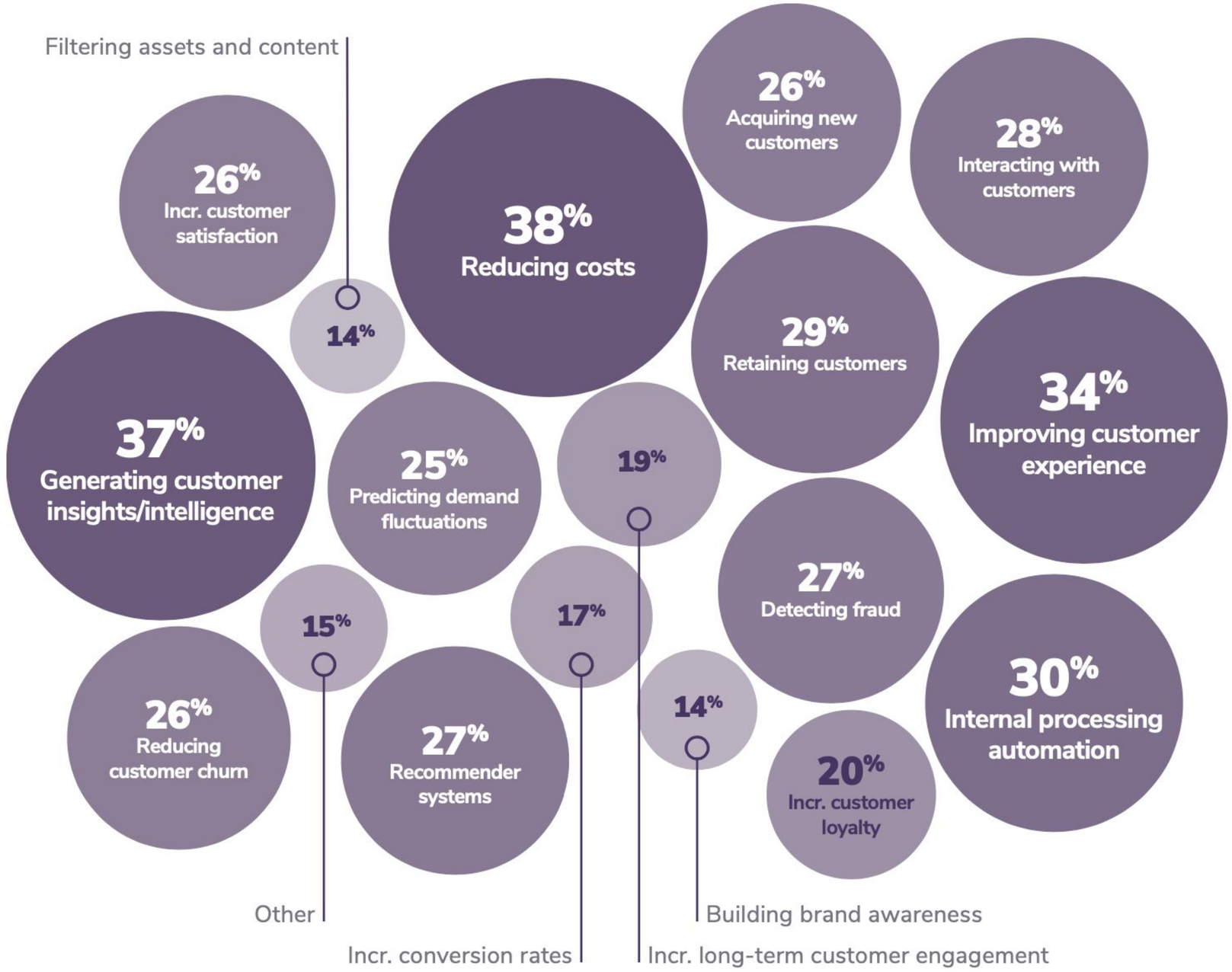
Decision Making

Customer Service Level Improvement

Customer Retention Program

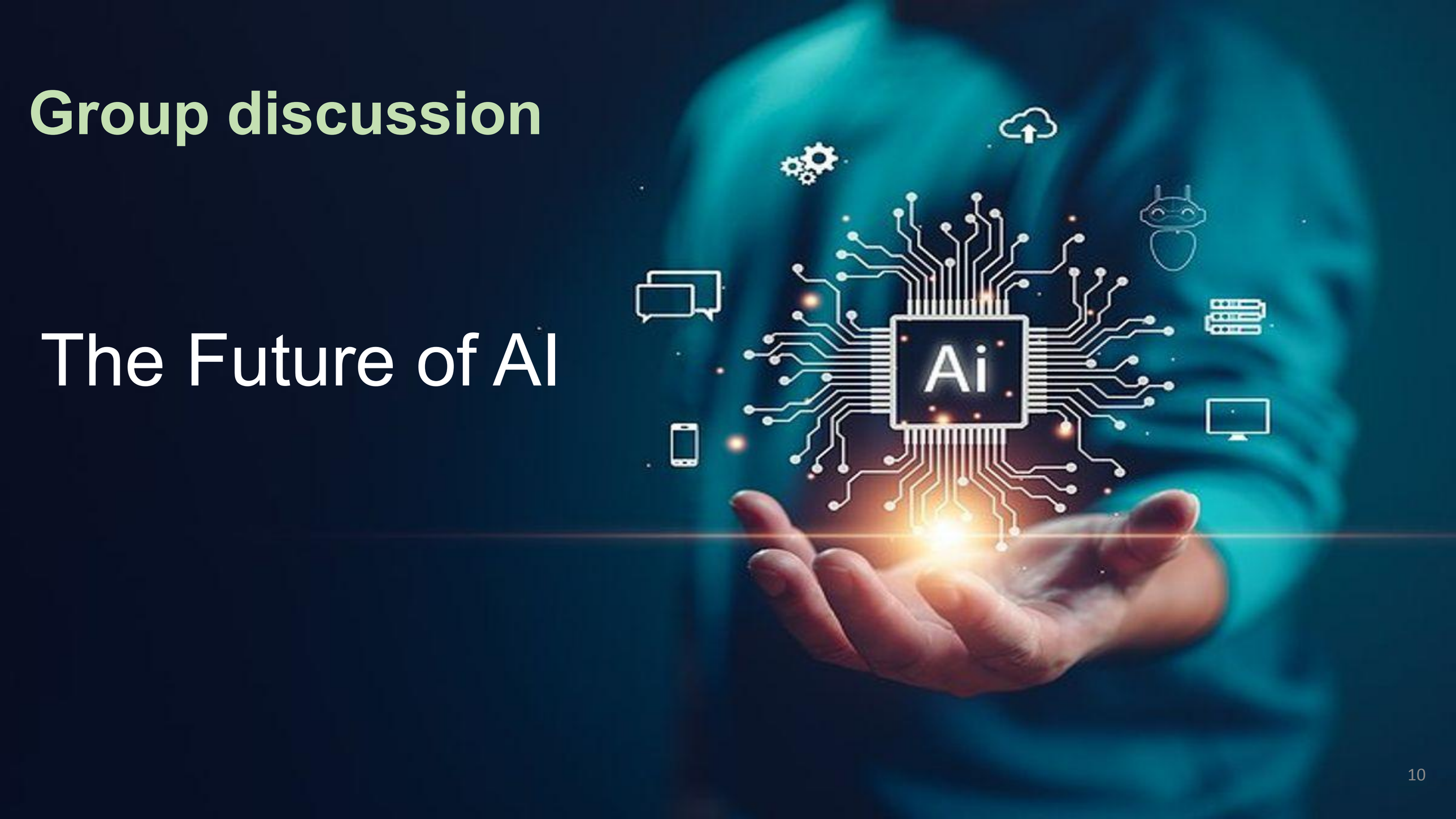
Marketing

Machine learning use case frequency

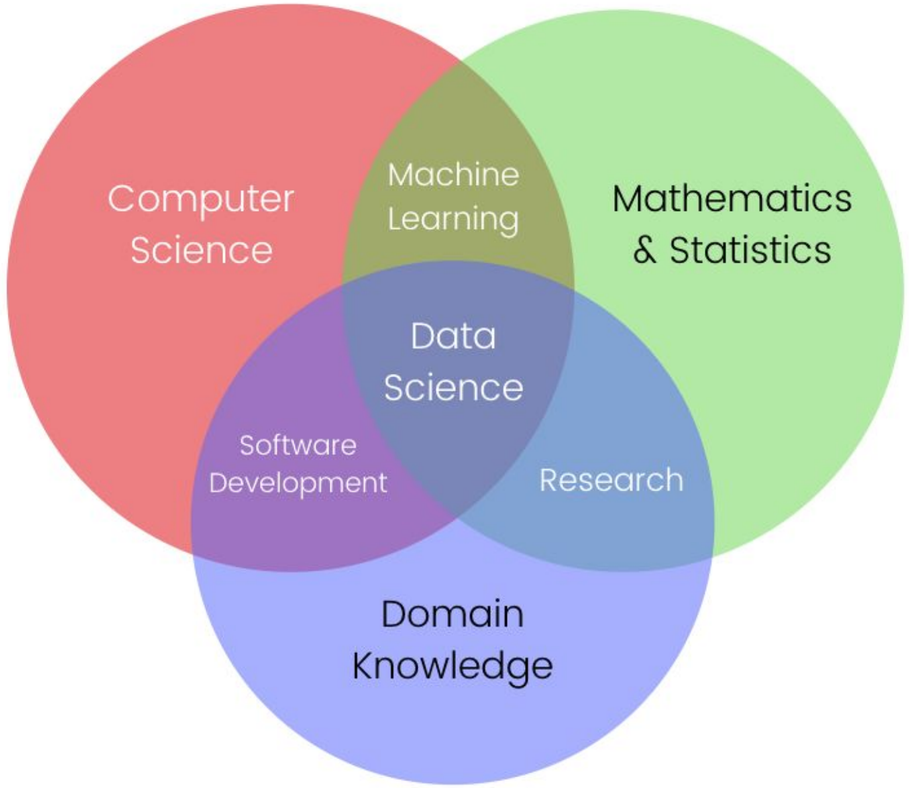


Group discussion

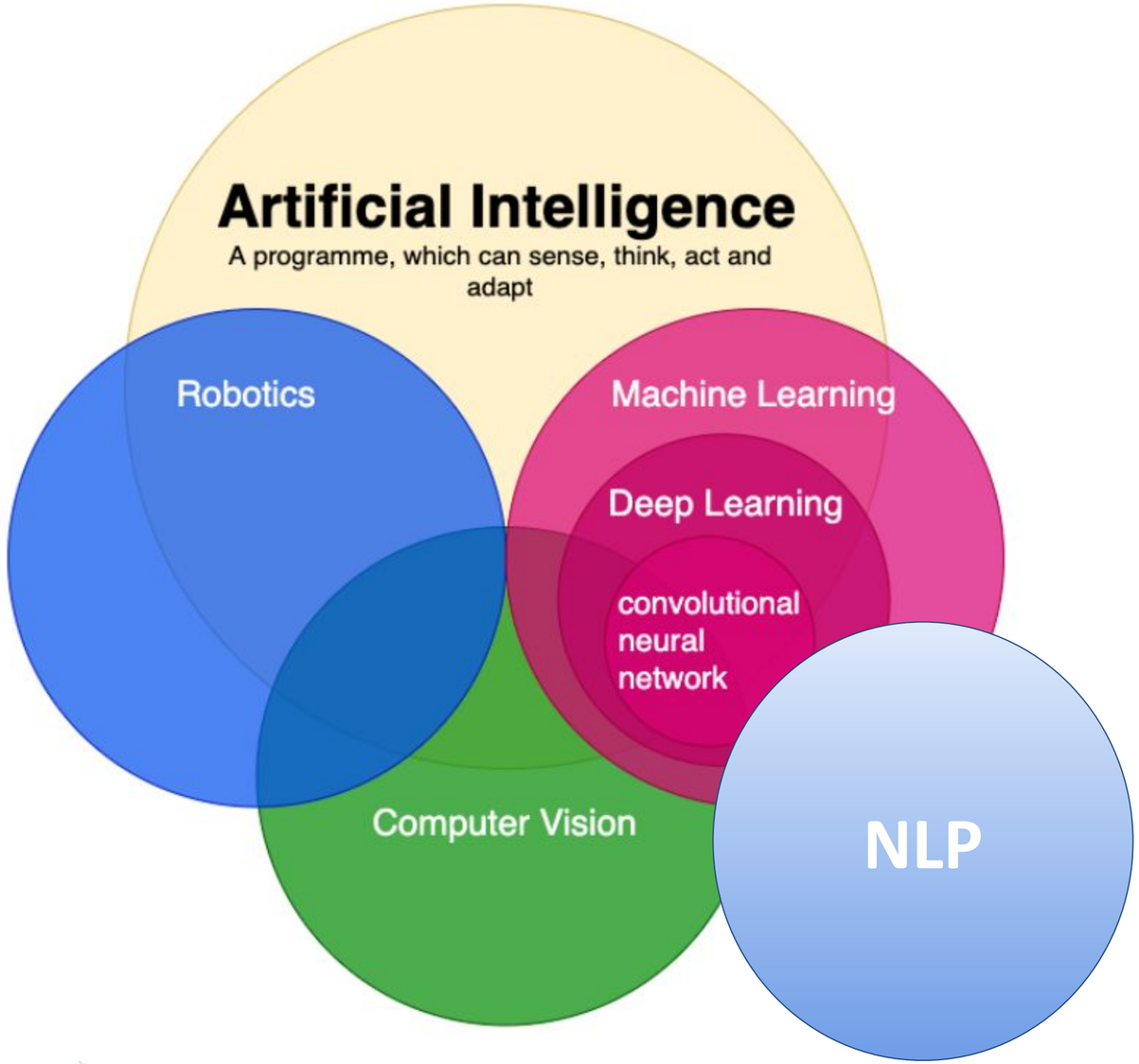
The Future of AI

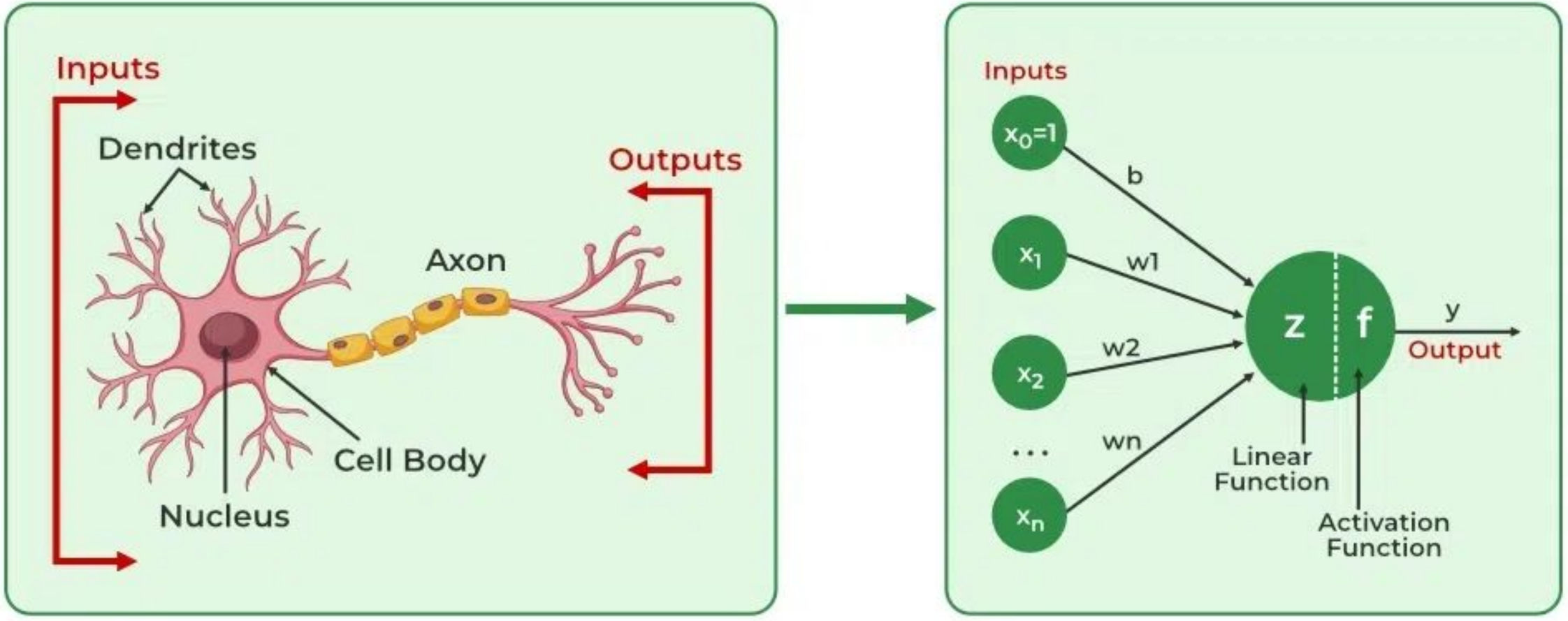


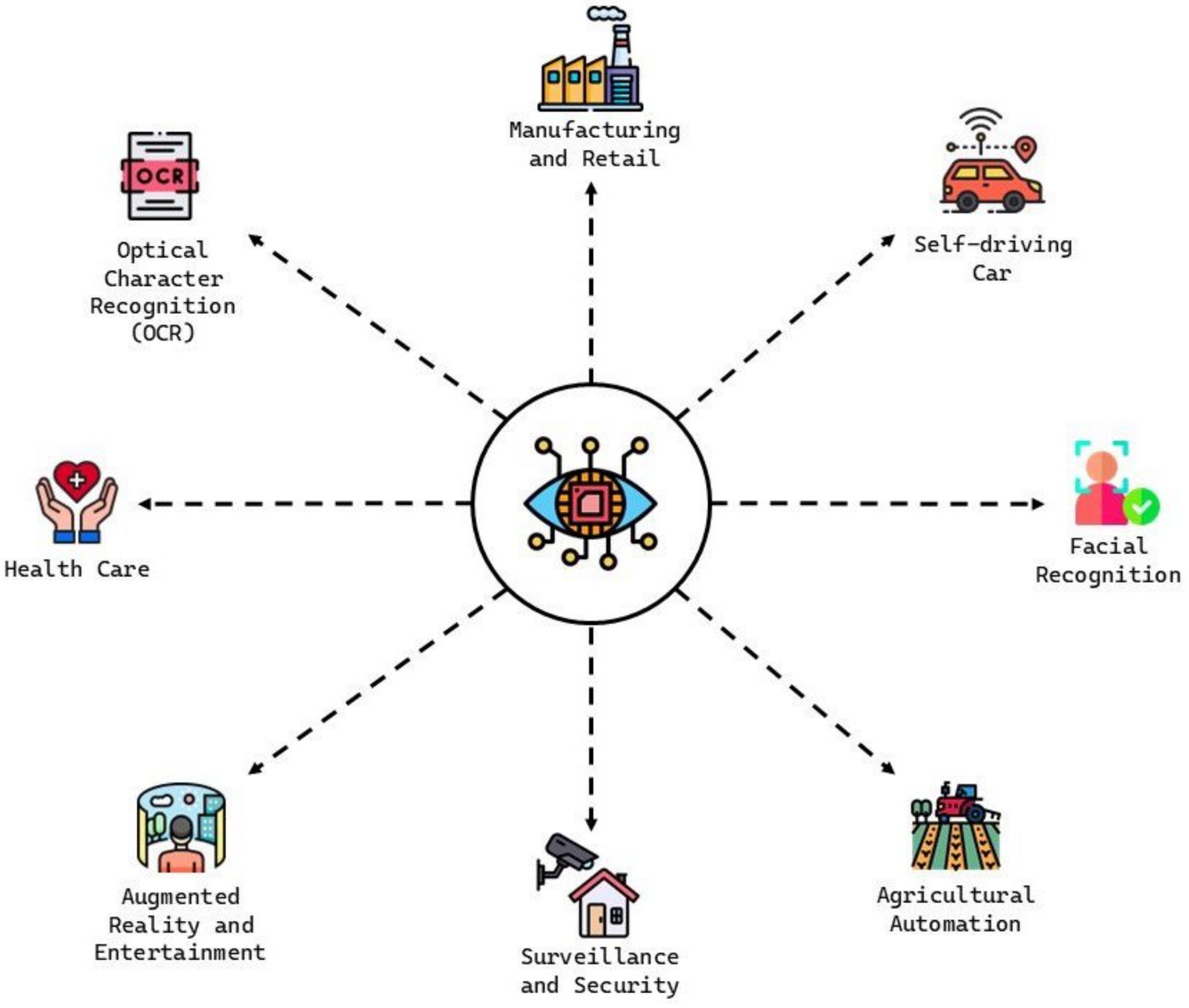
Understanding Data Science

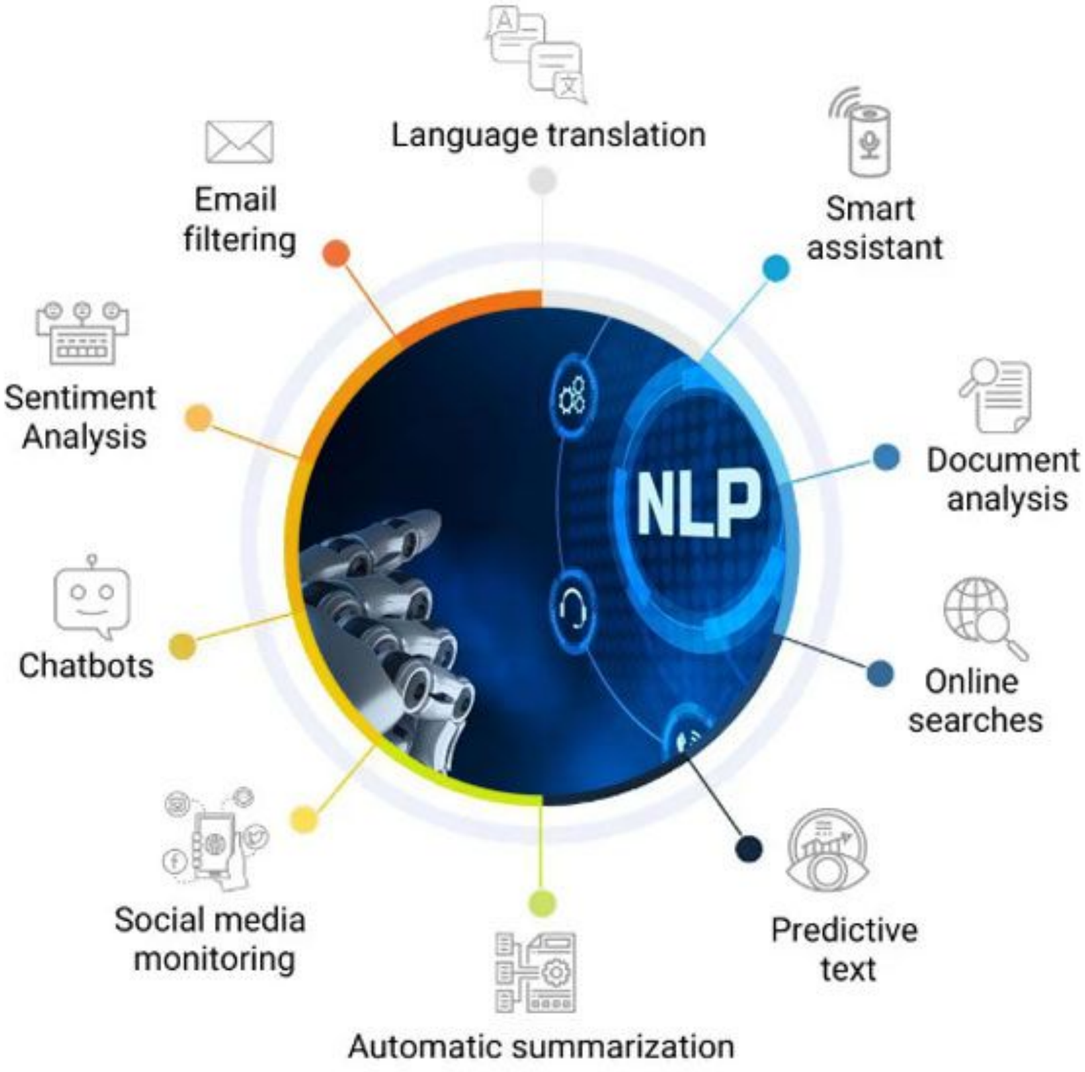


8:20 Break


Artificial Intelligence
A programme, which can sense, think, act and adapt
Robotics
Machine Learning
Deep Learning
convolutional neural network
Computer Vision
NLP


Inputs
Dendrites
Outputs
Axon
Cell Body
Nucleus
Inputs
$x_0=1$
b
x_1
w1
x_2
w2
...
wn
x_n
z
f
y
Output
Linear Function
Activation Function


Manufacturing and Retail
Optical Character Recognition (OCR)
Self-driving Car
Health Care
Facial Recognition
Augmented Reality and Entertainment
Surveillance and Security
Agricultural Automation


Language translation
Email filtering
Smart assistant
Sentiment Analysis
NLP
Document analysis
Chatbots
Online searches
Social media monitoring
Predictive text
Automatic summarization

Understanding Data Science

- Data is transformed into compelling narratives through storytelling.
- These insights drive strategic decision-making for organizations.
- It encompasses extracting and analyzing data in structured and unstructured forms.

The Essence of Data Science

- Data science explores, manipulates, and analyzes data to find answers.
- Just as other sciences study specific subjects, data science focuses on understanding data.
- Today’s world offers an abundance of data, algorithms, and accessible tools.
- The affordability and accessibility of these resources make data science more relevant than ever.

رواد مصر الرقمية

وزارة الاتصـــــالات
وتكنولوجيا المعلومات

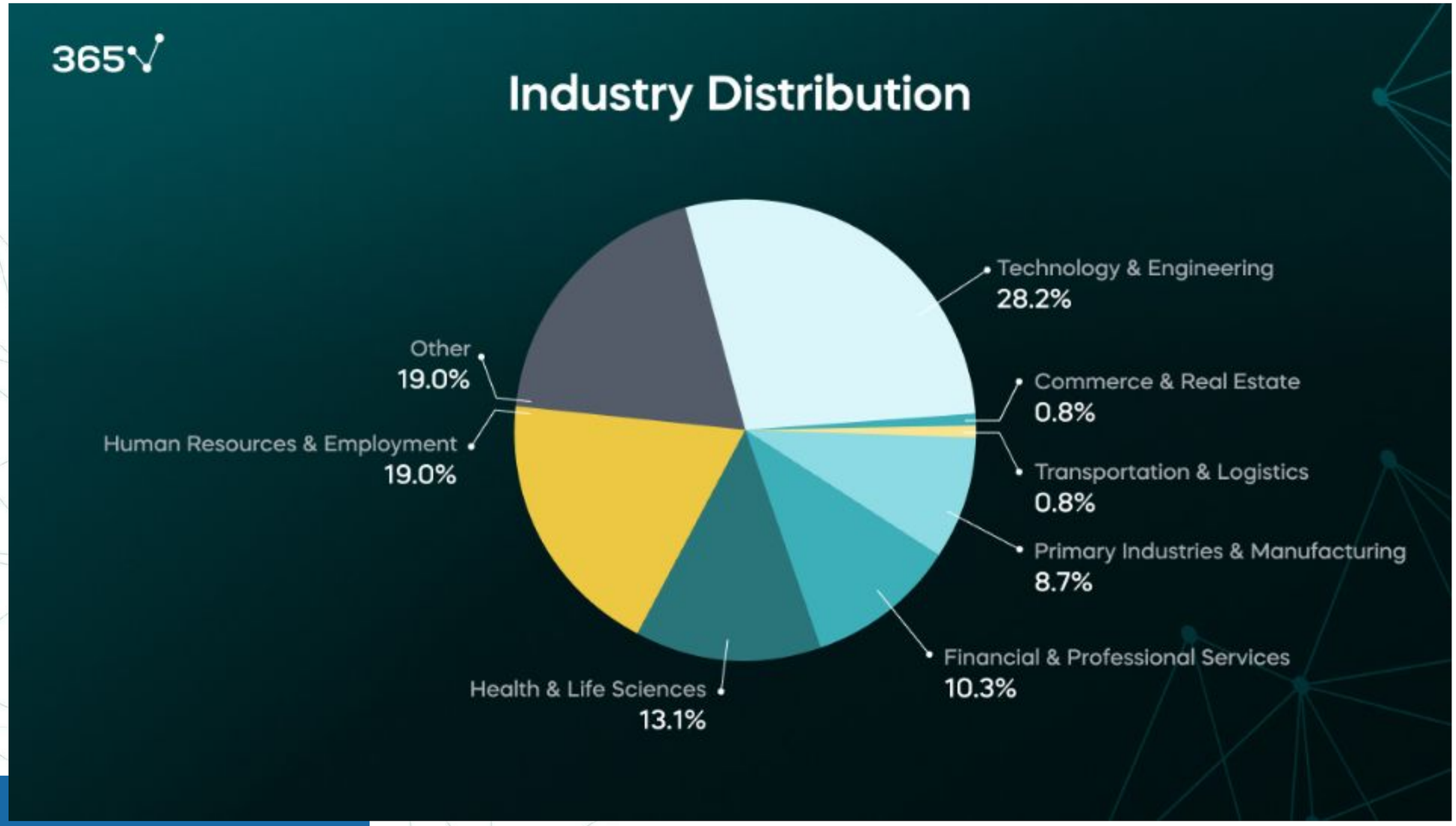
Fundamentals & Paths to a Career of Data Science

1. **Learn Programming**: Start with Python or R.
2. **Understand Math and Statistics**: Basics of linear algebra and statistics are essential.
3. **Data Understanding (Data Analysis, Visualization and Exploration)**
4. **Data Preprocessing and Cleaning**
5. **Machine Learning**: Learn basic algorithms like Linear Regression and Decision Trees.
6. **Real Projects**: Participate in projects or competitions like Kaggle.
7. **Continuous Learning**: Keep up with new courses and research.
8. **Choose a Specialization**: Focus on a specific field like financial analysis or healthcare
9. **Communication and Interpretation.**

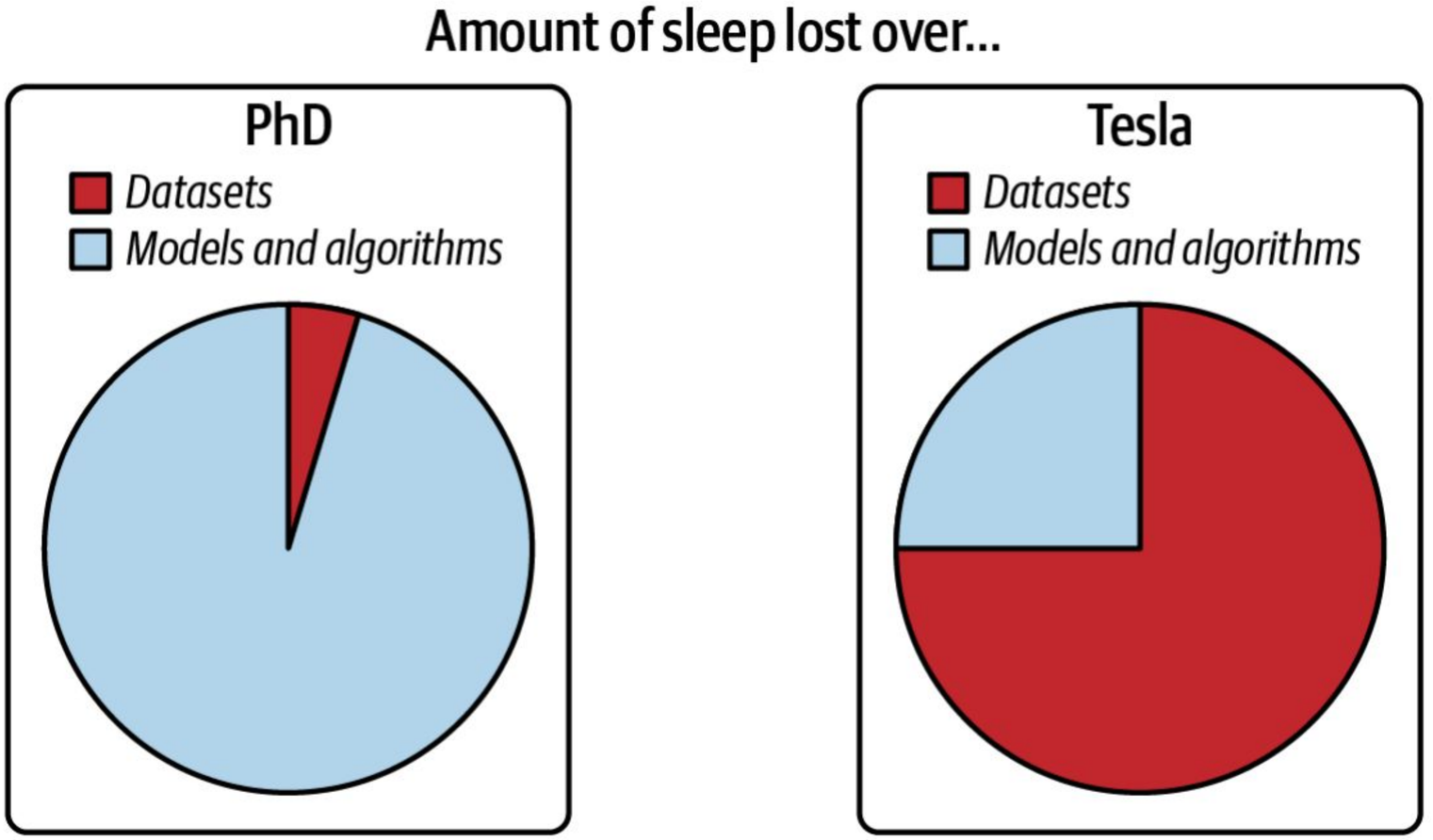
4/28/2024

- The Data Scientist Job Market

The Data Scientist Job Market in 2024 [Research on 1,000 Job Postings]



Data in research vs. data in production



Andrej Karpathy (Former Director of AI at Tesla)

AI in production: expectation

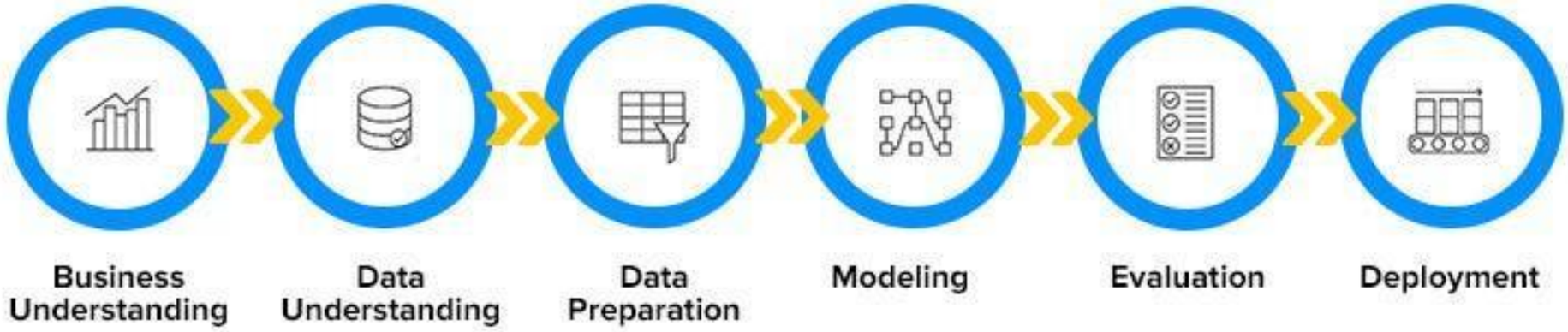
1. Collect data
2. Work in Data
3. Train model
4. Deploy model
5. 

AI in production: reality

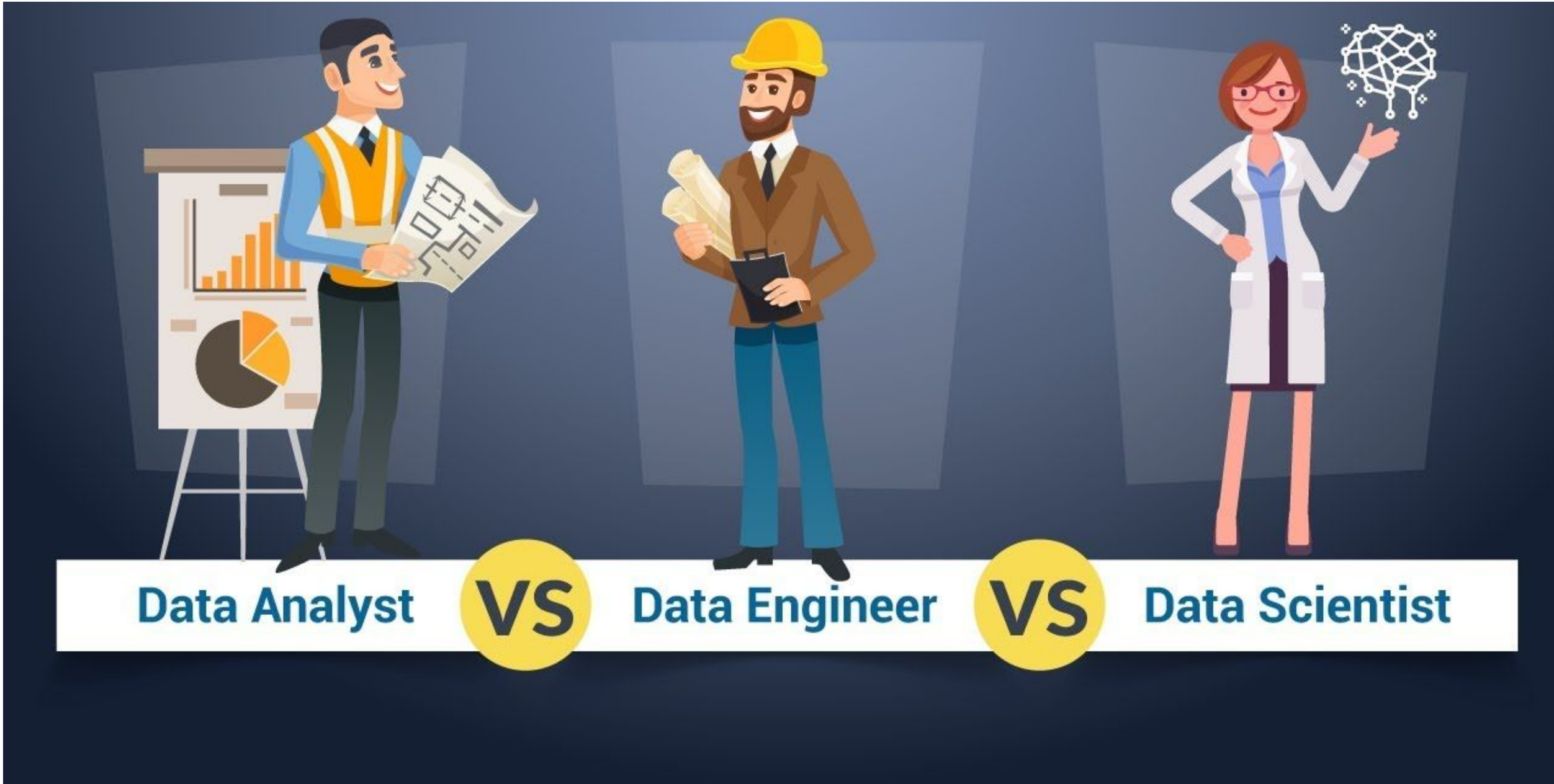
1. Choose a metric to optimize
2. Collect data
3. Work in Data
4. Train model
5. Realize many labels are wrong -> relabel data
6. Train model
7. Model performs poorly on one class -> collect more data for that class
8. Train model
9. Model performs poorly on most recent data -> collect more recent data
10. Train model
11. Deploy model
12. Dream about $$$
13. Wake up at 2am to complaints that model biases against one group -> revert to older version
14. Get more data, train more, do more testing
15. Deploy model
16. Pray
17. Model performs well but revenue decreasing
18. Cry
19. Choose a different metric
20. Start over

Step 16 and 18 are essential

The Data Science Process



Data Different Jobs



Data Scientist VS **Data Analyst** VS **Data Engineer** VS **Machine Learning Engineer**

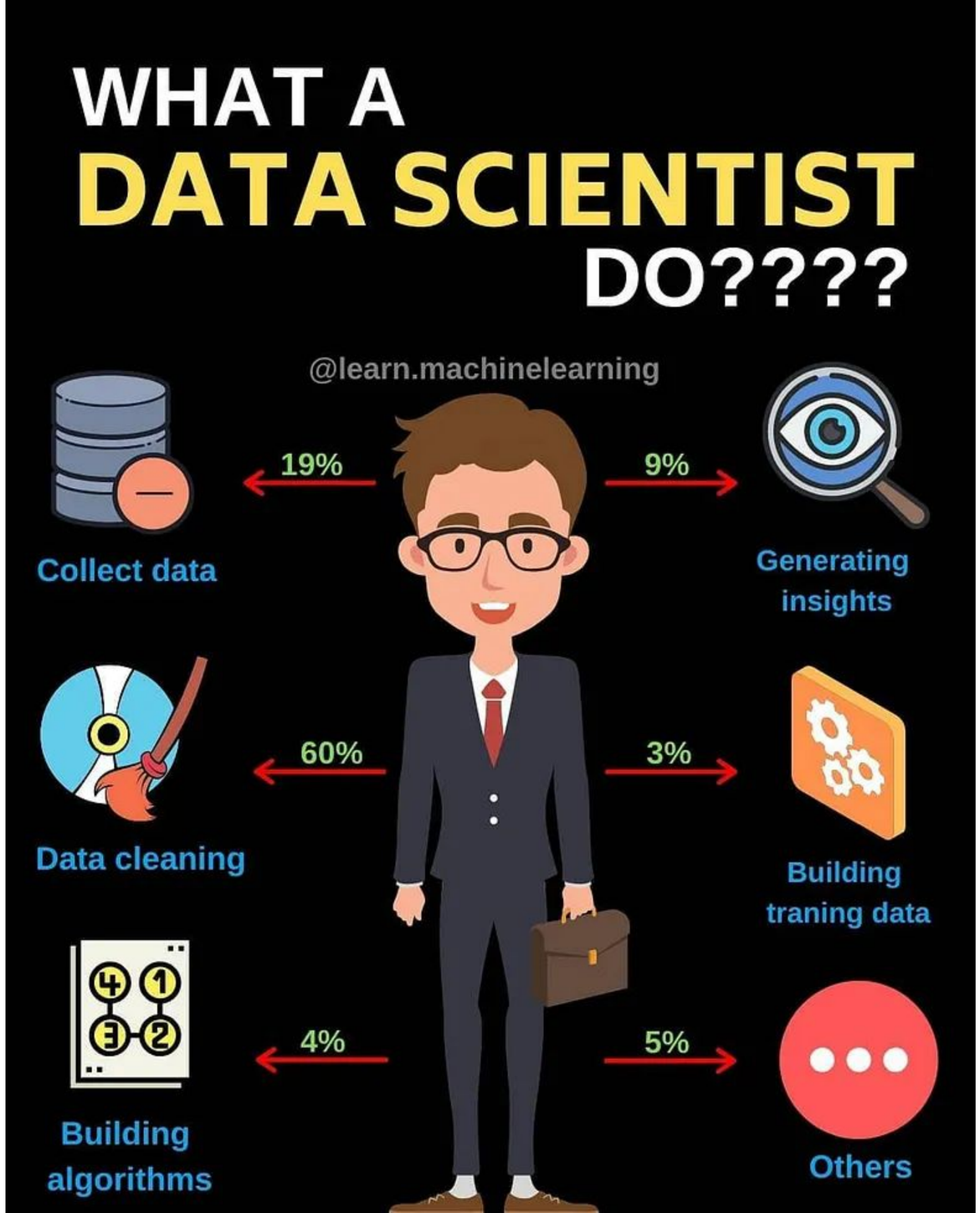
- **Data Analyst:**
  - Think of a Data Analyst as a detective who is always investigating, "What happened, and how can we improve things?"

- **Data Engineer:**
  - Think of a Data Engineer as the engineer who builds the pipelines that deliver clean water (data) to everyone.

- **Data Scientist:**
  - Think of a Data Scientist as the explorer guiding the ship into the unknown, always asking, "What can I discover in this data?"

- **Machine Learning Engineer:**
  - Think of a Machine Learning Engineer as the engineer who turns theoretical designs into reality, making machines work intelligently in the real world.

What Data Scientists Do



Ministry of Communications and Information Technology

My Advice for New Data Scientists

Digital Egypt Pioneers

1. Use curiosity to fuel problem-solving
2. Master the basics
3. Work on real projects
4. Embrace failure
5. Learn to tell the story
6. Never stop learning
7. Build your network
8. Be adaptable
9. Focus on quality, not just quantity
10. Always think about added value
11. Be solution-oriented, not tool-oriented
12. Share your work and learn from others

Data Scientist Skills

رواد مصر الرقمية

وزارة الاتصالات وتكنولوجيا المعلومات

01

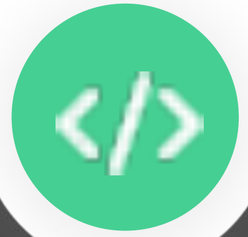
Data Analysis

Ability to analyze large datasets using statistical methods and machine learning algorithms.

02

Programming Skills

Proficiency in languages like Python, R, or SQL for data manipulation and analysis.



03

Data Visualization

Creating visual representations of data to communicate insights effectively.

04

Domain Knowledge

Understanding of the specific industry or domain to interpret data in context.

05

Problem-Solving

Applying analytical skills to solve complex business problems using data-driven approaches.

06

Communication

Effectively communicating findings and insights to stakeholders through reports and presentations.

TOP 50 AI & ML JOBS

Research & Development

- AI Research Scientist
- Machine Learning Researcher
- Deep Learning Scientist
- Computer Vision Research Scientist
- Natural Language Processing (NLP) Scientist
- AI Algorithm Engineer
- Reinforcement Learning Scientist

AI Implementation & Support

- AI Implementation Engineer
- AI Systems Integrator
- AI Support Specialist
- AI Deployment Engineer
- AI Technical Support Engineer

Business & Strategy

- AI Strategist
- AI Consultant
- AI Business Analyst
- AI Project Manager

AI Product Development

- AI Product Manager
- AI Product Owner
- Technical Product Manager (AI/ML Focus)
- AI Application Developer
- AI Solutions Architect

Niche AI Roles

- Robotics Engineer (AI Focus)
- Autonomous Systems Engineer
- AI Healthcare Specialist
- AI Financial Analyst
- AI Content Strategist

ML Engineering

- Machine Learning Engineer
- Machine Learning Infrastructure Engineer
- Deep Learning Engineer
- AI Software Developer
- Data Scientist (Machine Learning Focus)
- Machine Learning Operations (MLOps) Engineer
- AI/ML DevOps Engineer

AI Ethics & Governance

- AI Ethics Officer
- AI Governance Specialist
- AI Policy Advisor

Emerging Technologies

- Quantum Machine Learning Specialist
- AI Edge Computing Engineer
- AI for IoT (Internet of Things) Engineer
- Generative AI Specialist

Leadership & Executive

- Chief AI Officer (CAIO)
- AI Engineering Lead
- Head of AI/ML
- Director of AI Research

Data Roles

- Data Engineer
- Data Scientist
- Data Analyst
- Big Data Engineer
- Data Architect

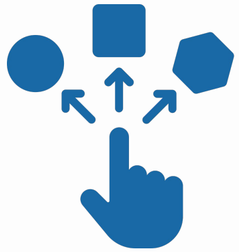
رواد مصر الرقمية

وزارة الاتصـــــالات
وتكنولوجيا المعلومات

Big Data

Volume

Dealing with large volumes of data that traditional systems cannot handle efficiently.



Variety

Handling diverse data types such as structured, semi-structured, and unstructured data.

Velocity

Processing data streams in real-time or near real-time to derive timely insights.

Big Data

Veracity

Ensuring data quality and reliability in a big data environment.

Value

Extracting actionable insights and value from big data for business decisions.

Tools & Technologies

Using platforms like Hadoop, Spark, and cloud services for big data processing and analytics.

Harnessing Big Data

- **Value:** Ability to derive value from data beyond profit, including social and personal benefits.
- **Examples of V's in Action:** YouTube uploads, world population's digital interactions, data types.
- **Coping with Big Data Challenges:** Traditional tools insufficient, alternative tools like Apache Spark and Hadoop.
- **Distributed Computing Power:** Tools enable extraction, loading, analysis, and processing of massive data sets.
- **Enriching Services:** Insights from Big Data allow organizations to better connect with customers.
- **Personal Data Journey:** From devices to Big Data analysis, data impacts services and returns to users.

Understanding Different Types of File Formats

Understanding File Formats

File formats are the different structures and encoding methods used to store and organize data in a digital file. The choice of file format directly affects how data is stored, retrieved, and processed. Different file formats serve various purposes, from simple text storage to complex data structures used in machine learning and big data applications. Below are key reasons why understanding file formats is important:

- **Compatibility**: Ensures that the data can be read and processed across different programs and systems.
- **Efficiency**: The right file format can reduce storage space and increase processing speed.
- **Data Integrity**: Helps maintain the structure and accuracy of the data during transfer or sharing.

Choosing the appropriate file format is crucial for ensuring data is handled properly, whether you're working on small data sets or large-scale projects. Let's take a look at a few commonly used file formats in data science.

Overview of File Formats

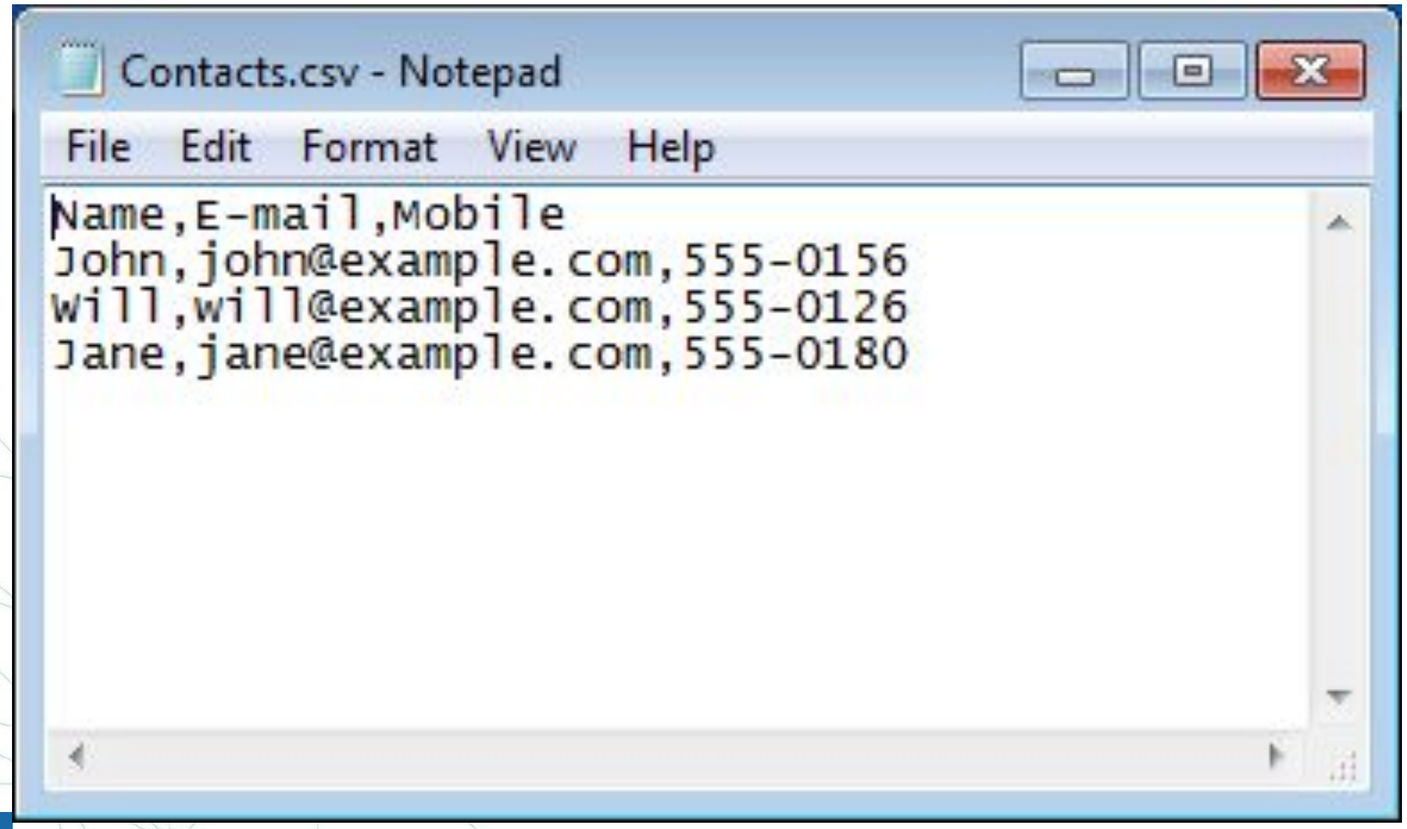
CSV (Comma-Separated Values)

- **Description**: A simple text-based format used to store tabular data where values are separated by commas.
- **Development**: An older, widely-used format for exchanging data between different systems.
- **Features**: Maintains a structured layout that's easy for both humans and machines to read, and can be opened in text editors or spreadsheet programs like Excel.
- **Advantages**: Lightweight and easy to handle, ideal for structured data like tables.
- **Common Use**: Used for storing organized data such as sales records, survey results, or data transferred between databases and analytical tools.

رواد مصر الرقمية

Example of File Formats

The below image shows a CSV file which is opened in Notepad.



Overview of File Formats

JSON (JavaScript Object Notation)

- **Description**: A lightweight, text-based format used for representing structured data as key-value pairs.
- **Development**: Originally derived from JavaScript, now widely adopted across different platforms and languages.
- **Features**: Human-readable and machine-parsable, supports hierarchical and nested data structures.
- **Advantages**: Flexible, easy to work with, especially for APIs and data exchange between web applications.
- **Common Use**: Commonly used for transmitting data between a server and a web application, and for storing complex structured data like user profiles, configurations, and API responses.

Overview of File Formats

XLSX:

- Microsoft Excel's Open XML format.
- Data is organized into worksheets (rows and columns).
- Supports formulas, charts, and data manipulation.
- Widely used for data analysis and reporting.

XML:

- A markup language designed to encode data.
- Both human- and machine-readable.
- Often used for data exchange between systems and applications.
- Flexible and customizable structure for storing and transmitting data.

Overview of File Formats

- **PDF:**
- Developed by Adobe for consistent document presentation.
- Maintains formatting across different devices and platforms.
- Ideal for sharing documents where layout, fonts, and images must remain intact.
- Commonly used for reports, contracts, and presentations.

Example of File Formats

CSV

	A	B	C	D
1	ID	Gender	City	Monthly_I
2	ID000002C	Female	Delhi	20000
3	ID000004E	Male	Mumbai	35000
4	ID000007H	Male	Panchkula	22500
5	ID000008I	Male	Saharsa	35000
6	ID000009J	Male	Bengaluru	100000
7	ID000010K	Male	Bengaluru	45000
8	ID000011L	Female	Sindhudur	70000
9	ID000012M	Male	Bengaluru	20000
10	ID000013N	Male	Kochi	75000
11	ID000014O	Female	Mumbai	30000
12	ID000016O	Male	Mumbai	25000
13	ID000018S	Female	Surat	25000
14	ID000019T	Female	Pune	24000
15	ID000021V	Male	Bhubanes	27000
16	ID000022V	Female	Howrah	28000

JSON

```
{
   "Employee": [

      {

         "id":"1",

         "Name": "Ankit",

         "Sal": "1000",

      },
      {

         "id":"2",

         "Name": "Faizy",
```

```
<?xml version="1.0"?>

<contact-info>

<name>Ankit</name>

<company>Anlytics Vidhya</company>

<phone>+9187654321</phone>

</contact-info>
```

Q & A

Thank you!