

Data analyst vs Data Engineer Roadmap and more

Hessam Mohammad Hosseini



Data Skills on demand

FIGURE 1.2

Technology trends driving business transformation, 2025-2030

Share of employers surveyed that identify the stated technology trend as likely to drive business transformation

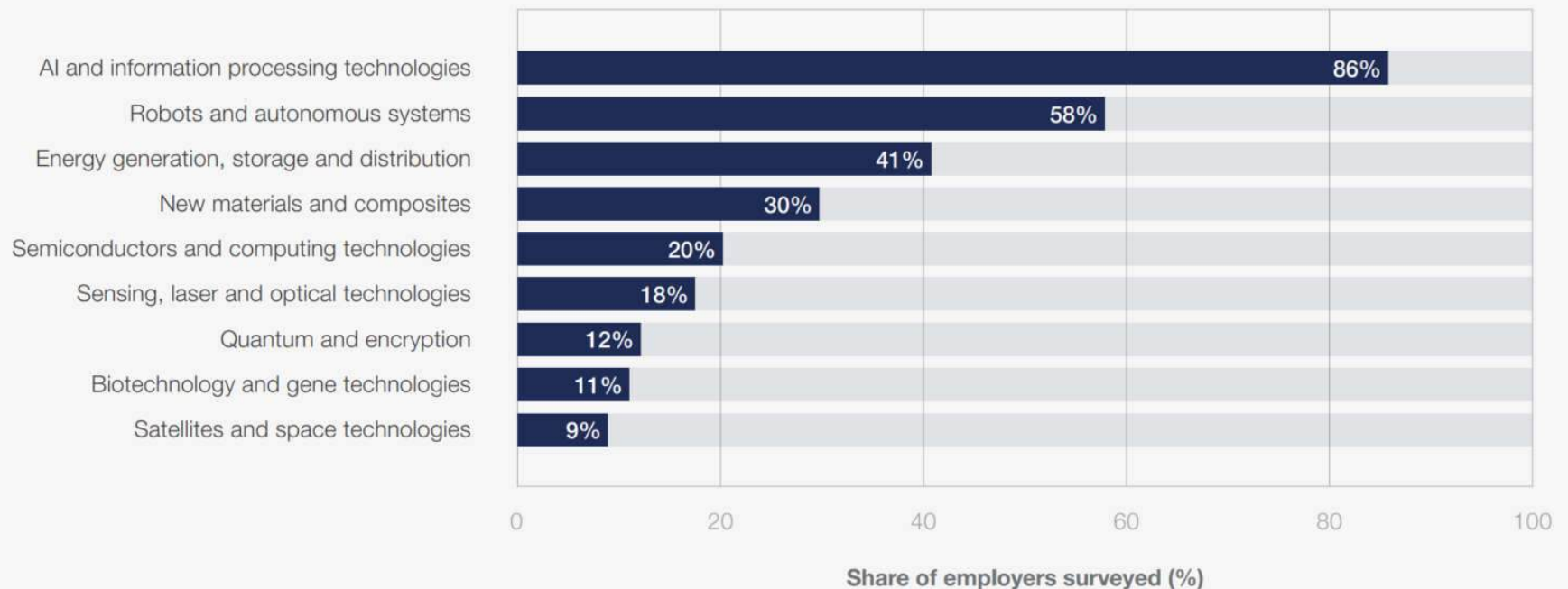


FIGURE 3.4 Skills on the rise, 2025-2030

Share of employers that consider skills to be increasing, decreasing, or remaining stable in importance. Skills are ranked based on net increase, which is the difference between the share of employers that consider a skill category to be increasing in use and those that consider it to be decreasing in use.

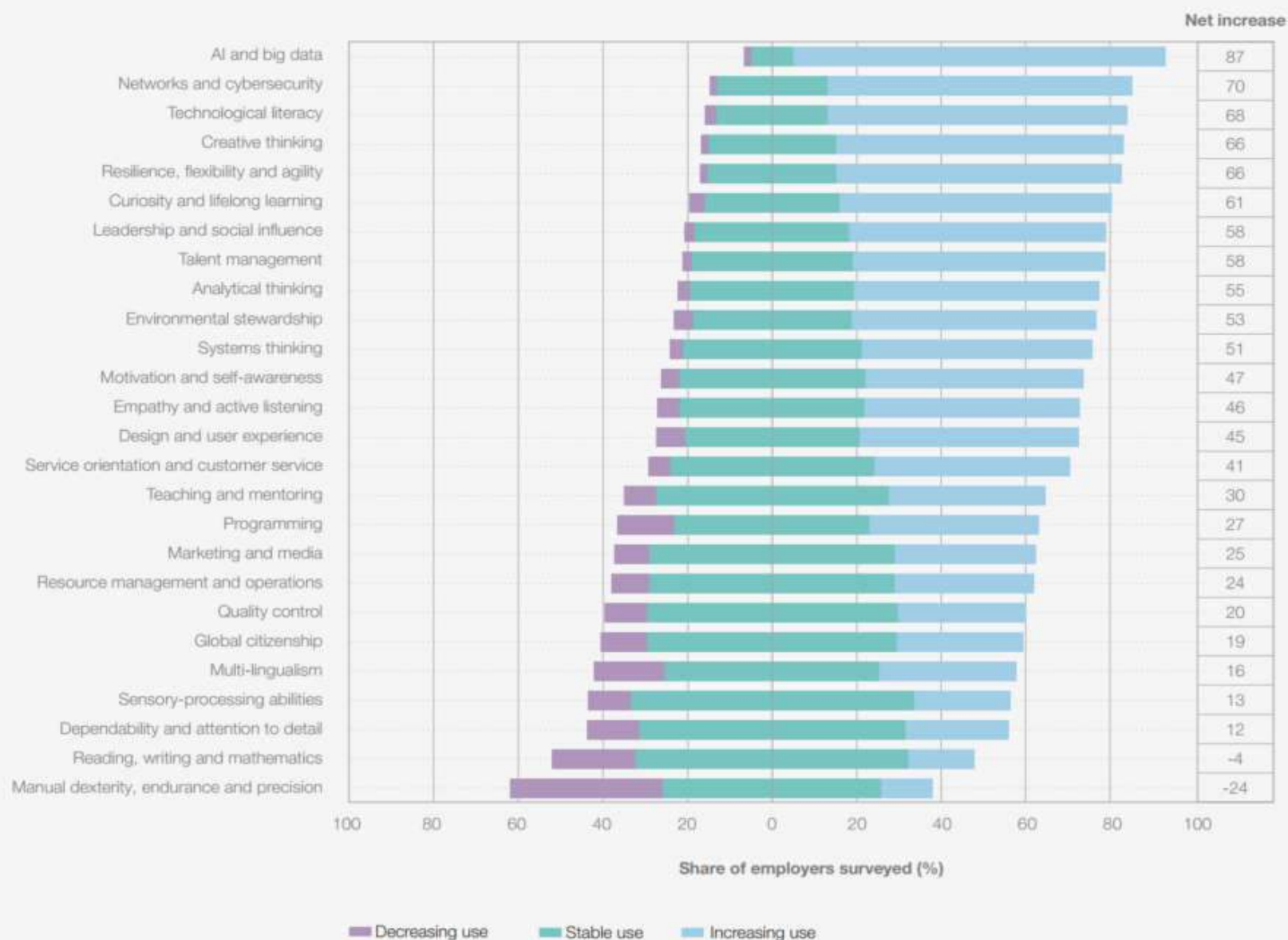
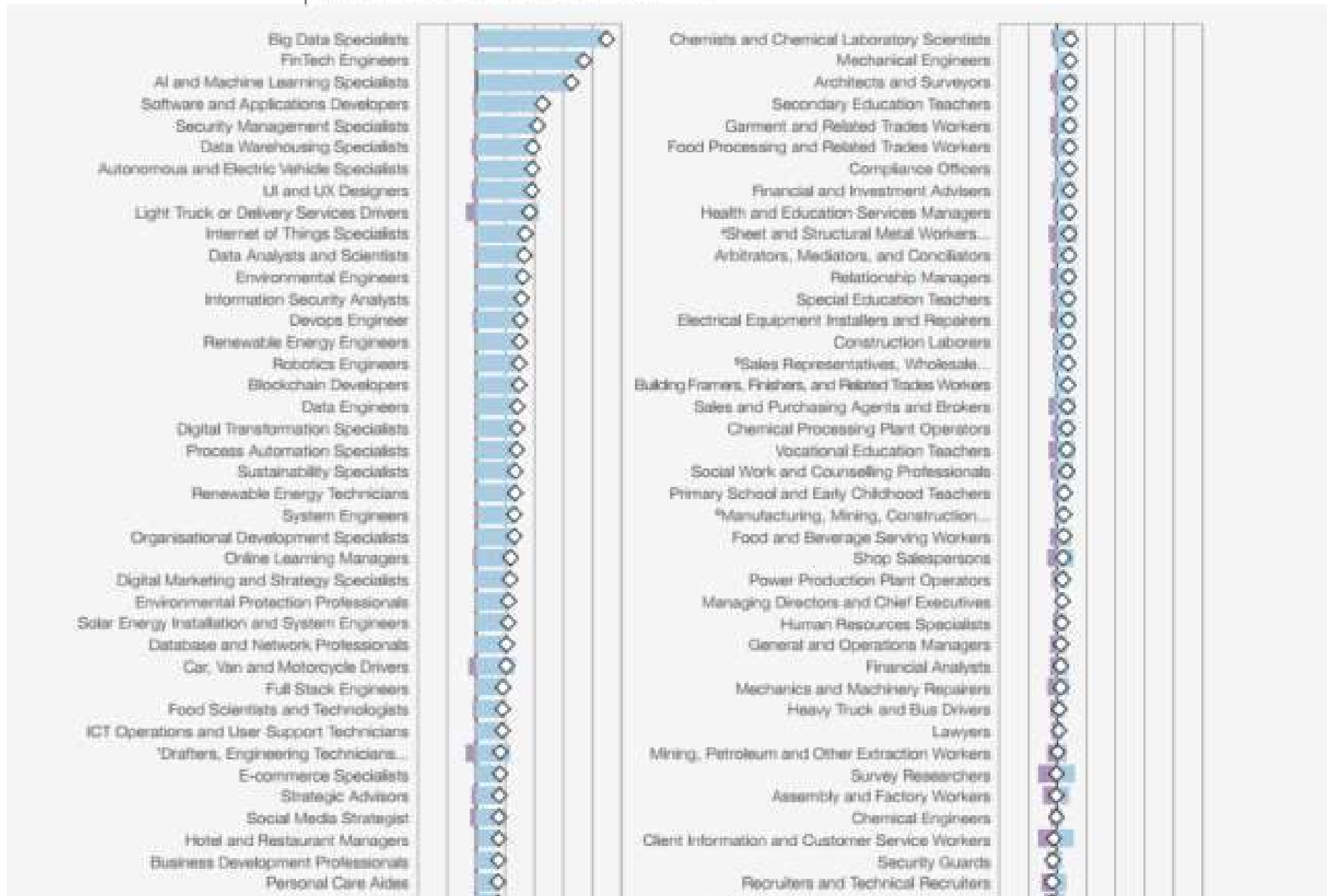


FIGURE 2.3 | **Job growth and decline (%), 2025-2030**

Projected job creation (blue) and displacement (purple) between 2025 and 2030, as a percentage of total current employment in the corresponding job role. The projected net growth or decline for each occupation over the next five years (diamonds) is calculated by subtracting total job displacement from total job creation.





■ Jobs created ■ Jobs displaced ◇ Net growth or decline

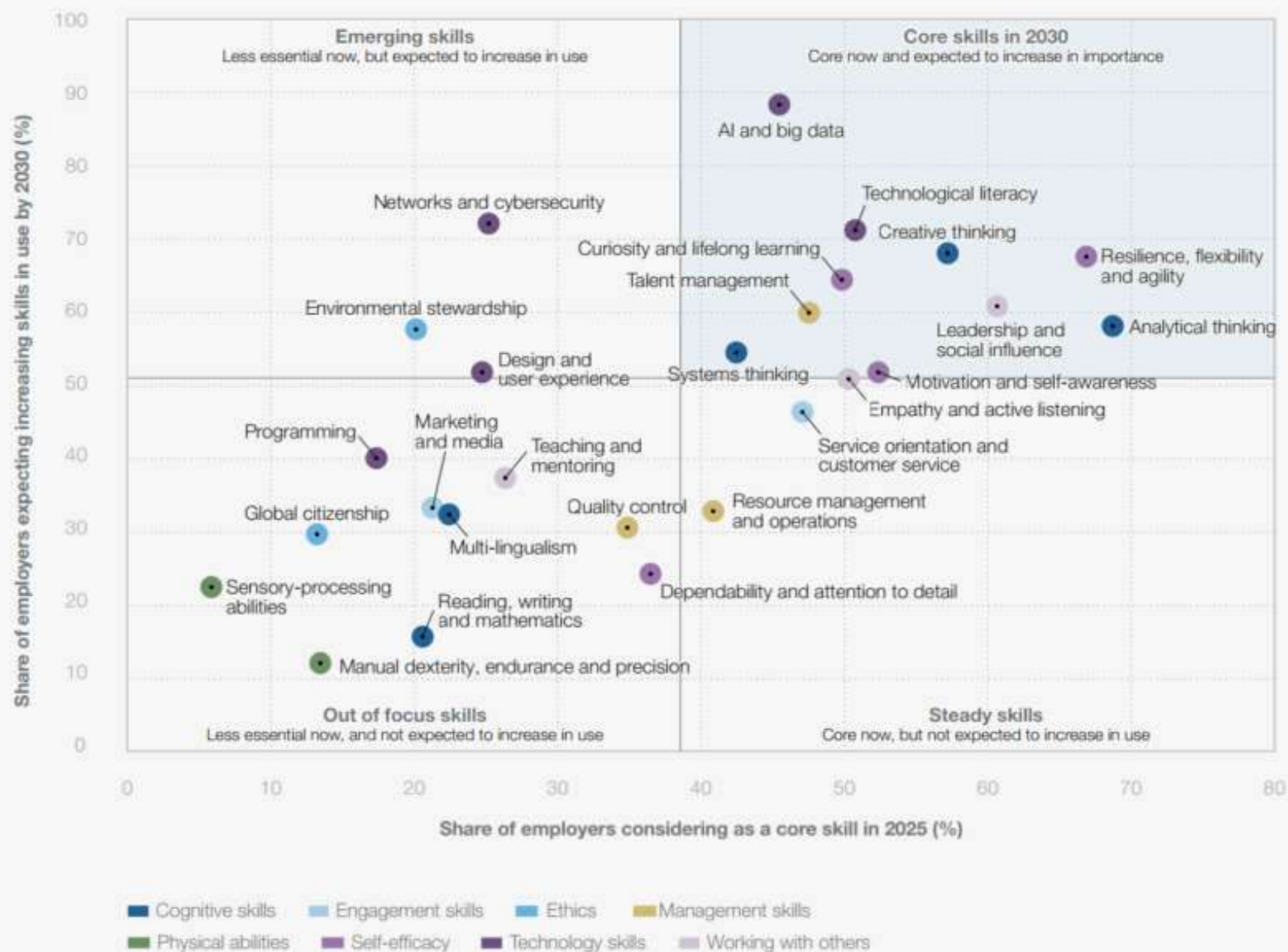
Note

*Crafters, Engineering Technicians, and Machine Technicians; *Farmworkers, Labourers, and Other Agricultural

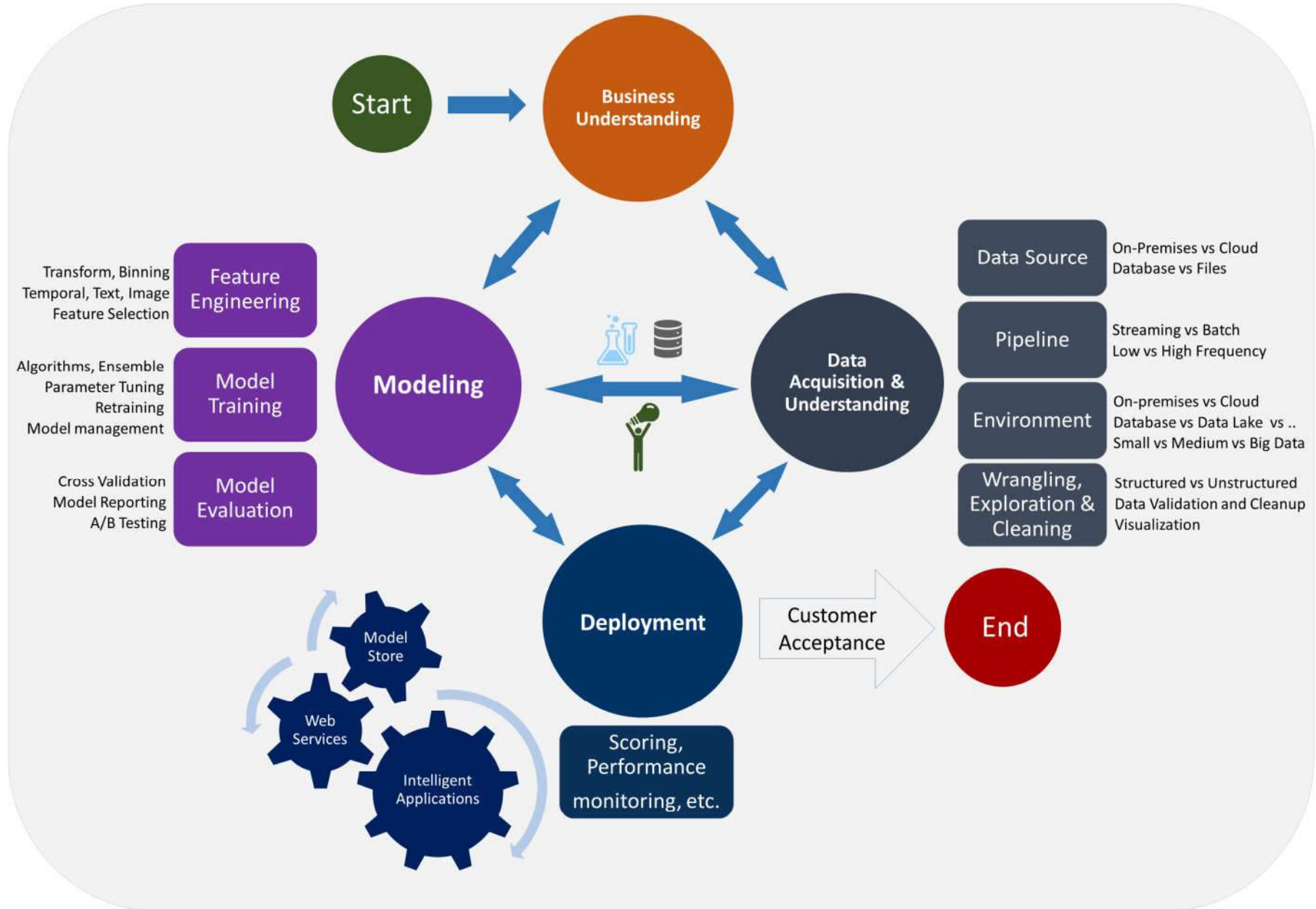
FIGURE 3.6

Core skills in 2030

Share of employers considering skills to be a core skill in 2025 and share of employers expecting skills to increase in importance by 2030.



Data Science Lifecycle



Data Science needs

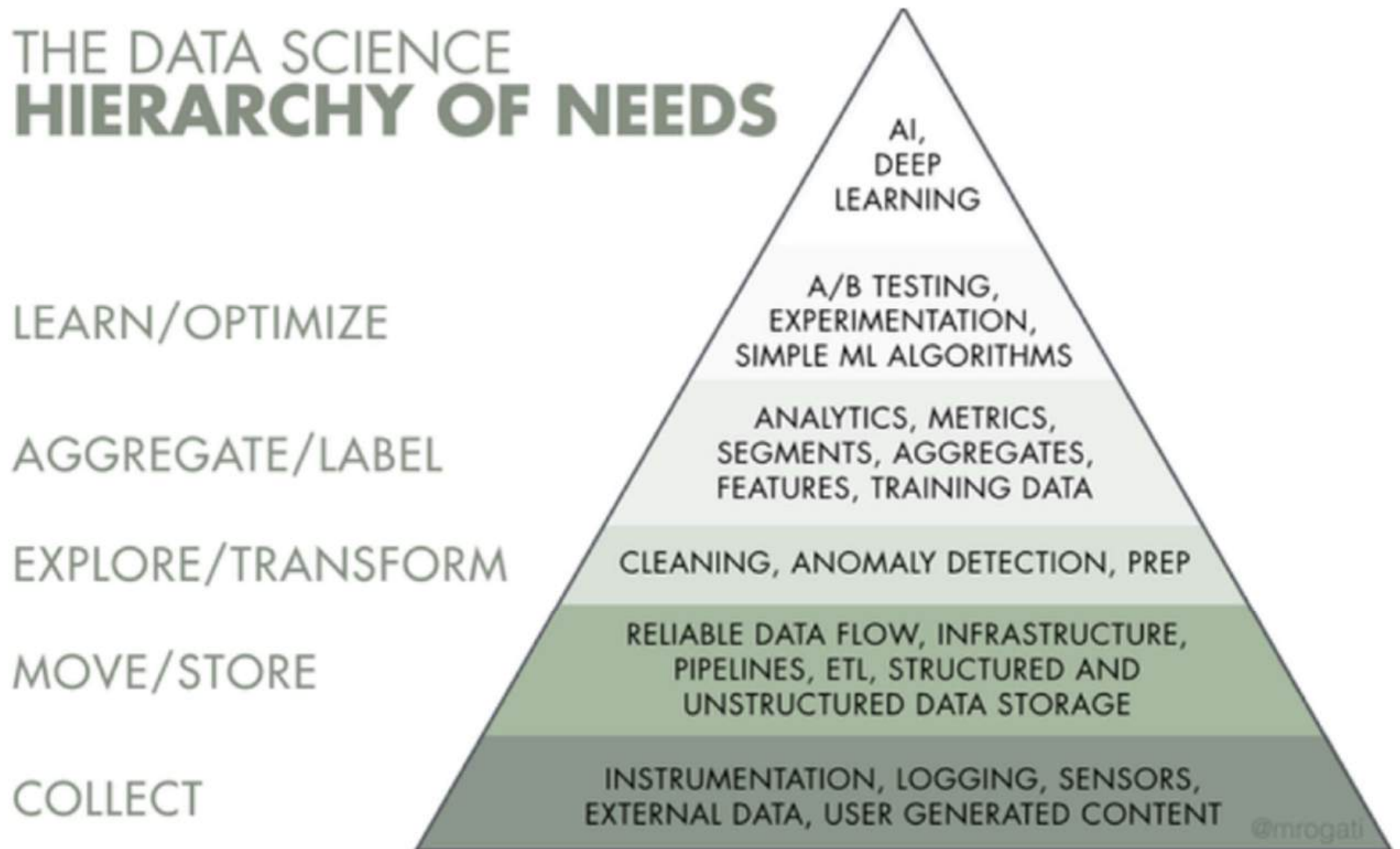
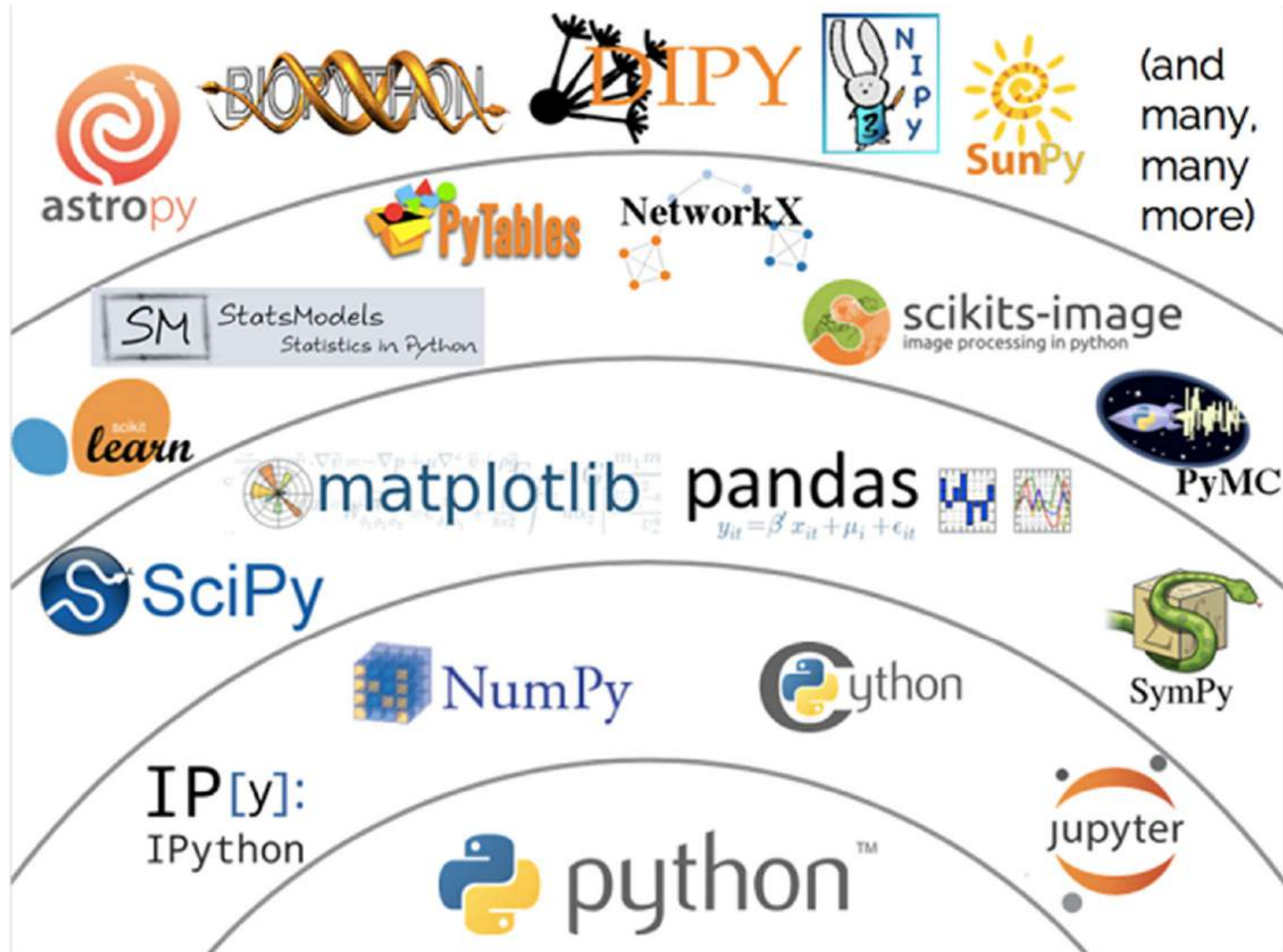


Figure 1-5. The Data Science Hierarchy of Needs

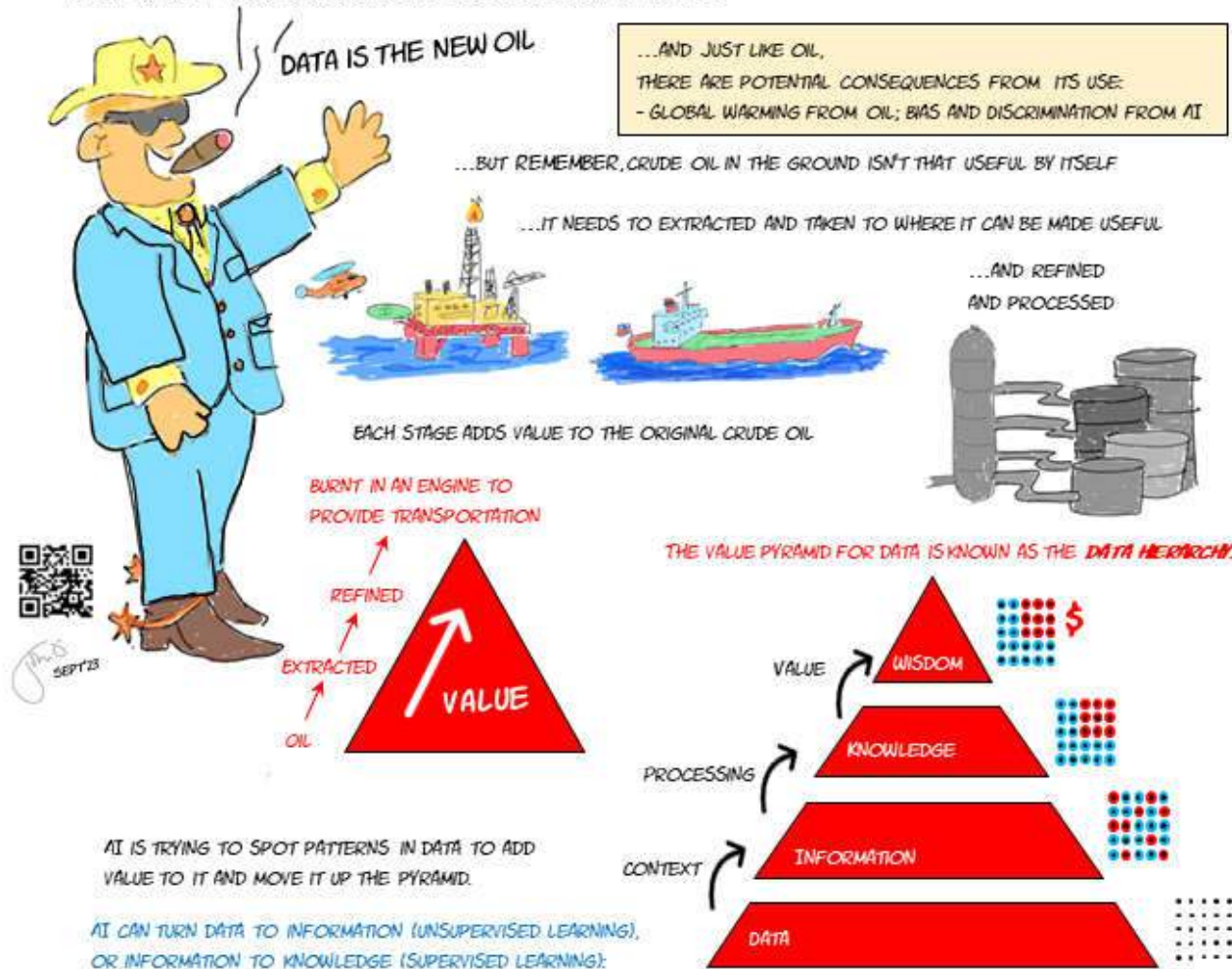
Tech stack



SKETCH 1, DATA HIERARCHY

AI CAN DO AMAZING THINGS LIKE LIVE LANGUAGE TRANSLATION AND IMAGE RECOGNITION - BUT THAT IS ONLY POSSIBLE BECAUSE OF THE DATA IT IS TRAINED UPON - AND HUGE AMOUNTS OF IT. IF YOU WANT TO KNOW ABOUT AI, YOU NEED TO KNOW ABOUT DATA.

DATA IS A COLLECTION OF FACTS OR VALUES, SUCH AS NUMBERS, WORDS, MEASUREMENTS, OBSERVATIONS, OR DESCRIPTIONS. TO BE USEFUL, IT NEEDS TO BE COLLECTED TOGETHER IN A SIMILAR FORM, SO THAT IT CAN BE ACCESSED, ANALYSED AND PATTERNS FOUND WITHIN IT- WHICH ALLOWS SOME MEANING TO BE DERIVED FROM IT.

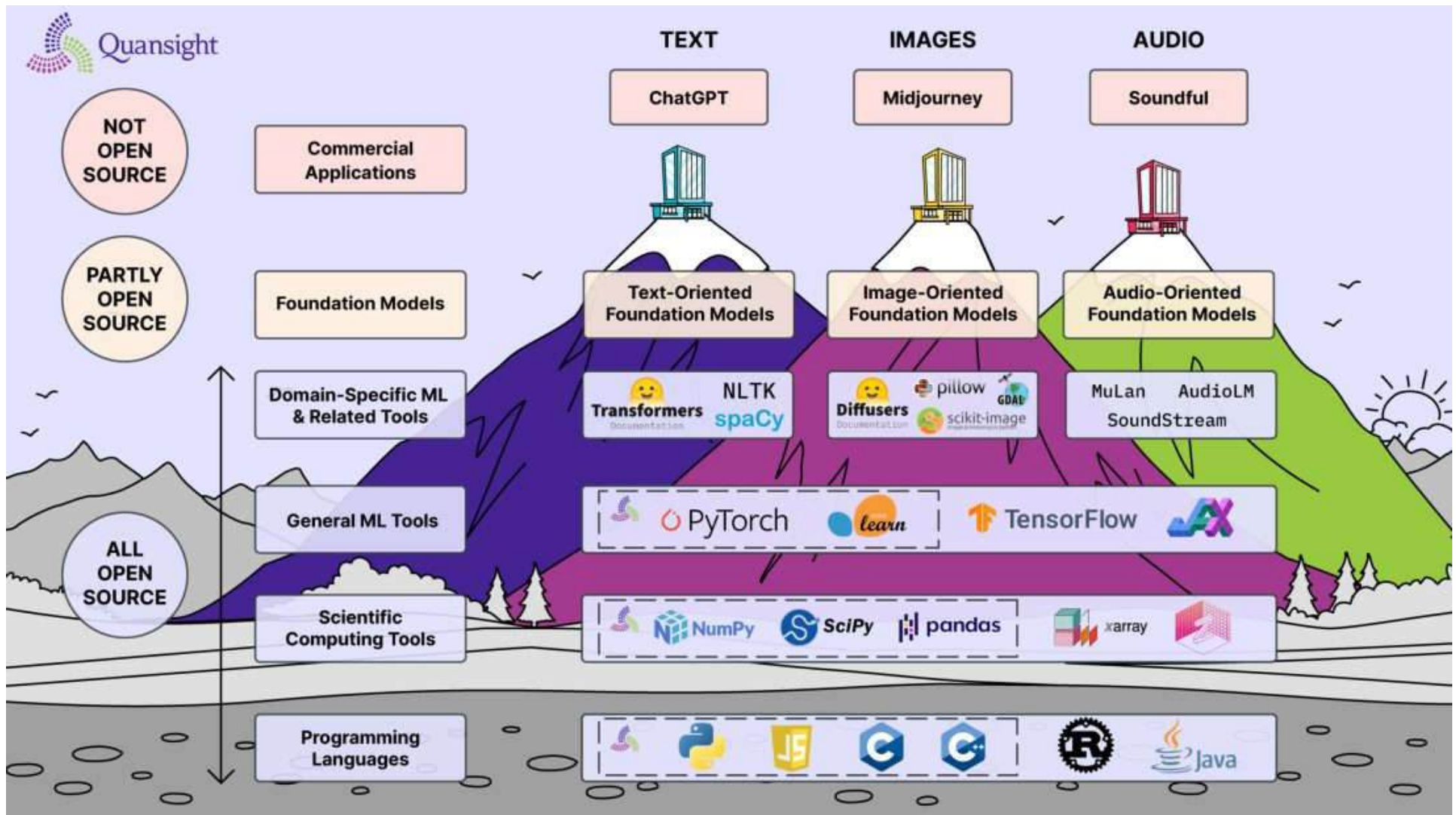


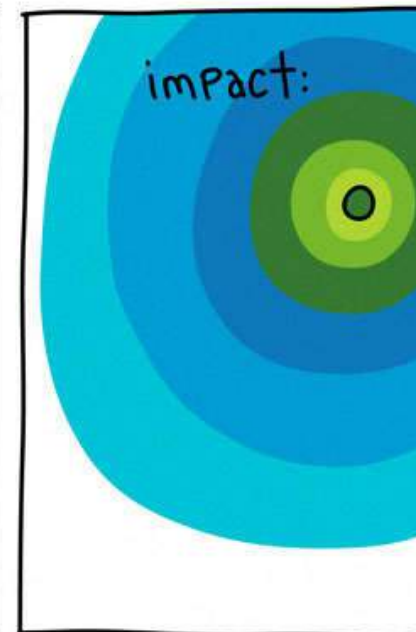
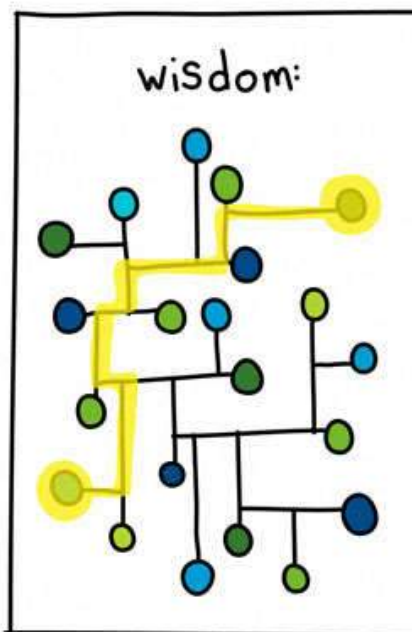
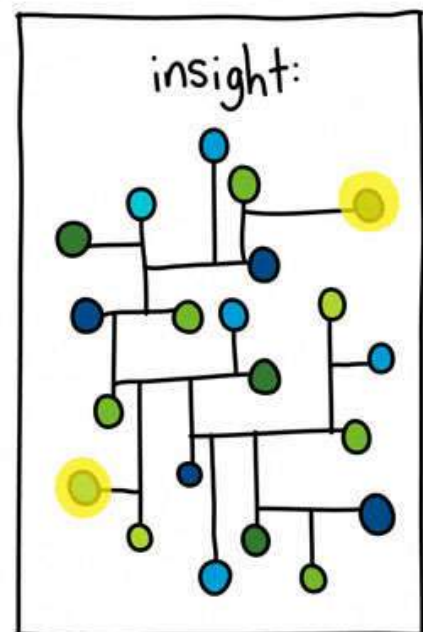
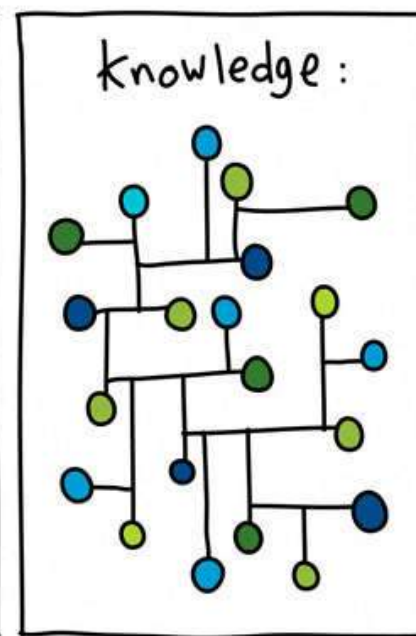
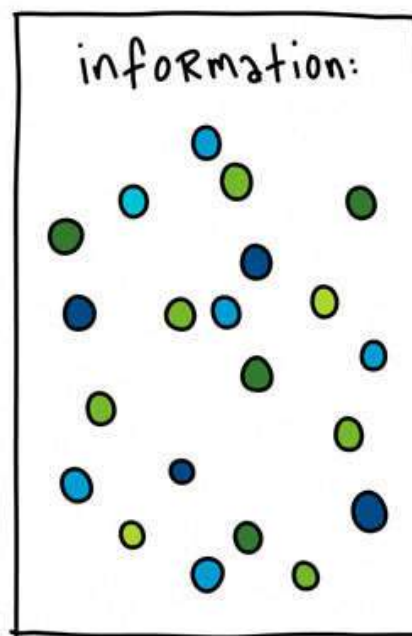
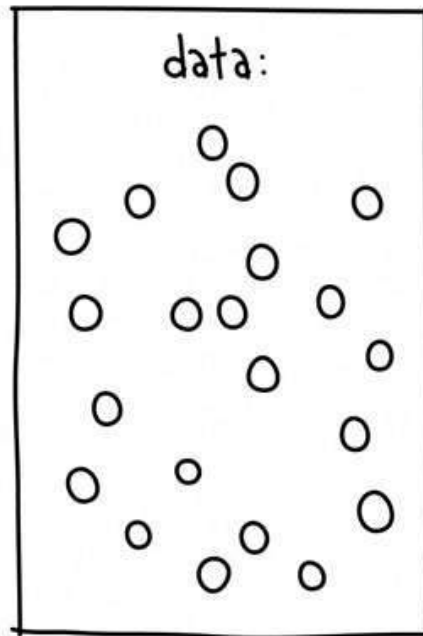
SPONSORED BY OUR FRIEND:

THEY HELP YOU USE AI TO...

ENJOYING THIS? THEN BE PART OF IT! SPONSOR THIS SKETCH - IDEAL FOR AI START-UPS
HELP OTHERS LEARN AND RAISE THE PROFILE OF YOUR BUSINESS AS THIS PAGE STARTS ITS VIRAL LEARNING JOURNEY

“If I have seen further, it is by standing on the shoulders of giants.”





Connecting

THE DATA SCIENCE HIERARCHY OF NEEDS

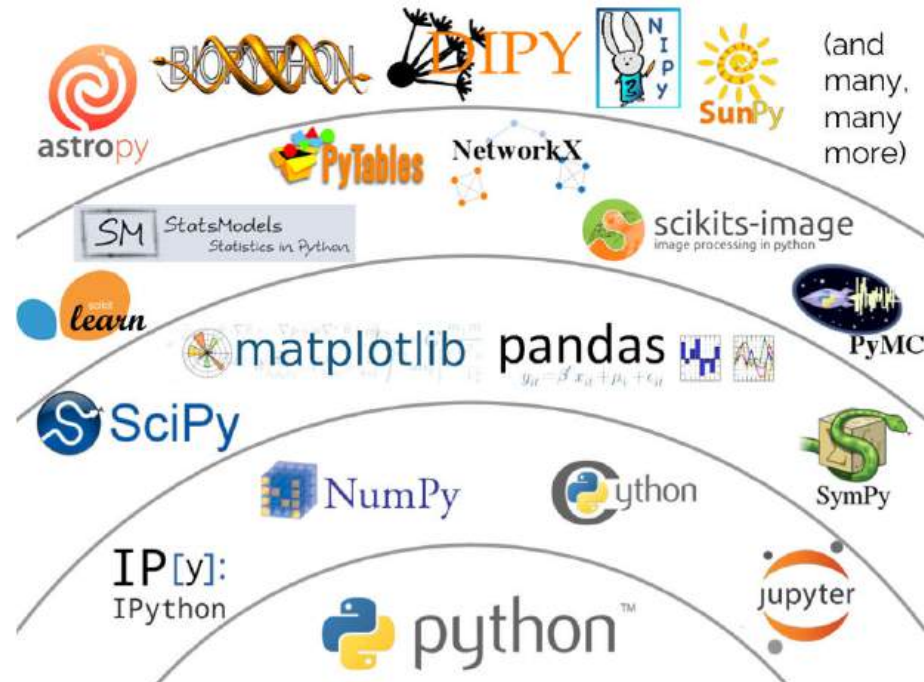
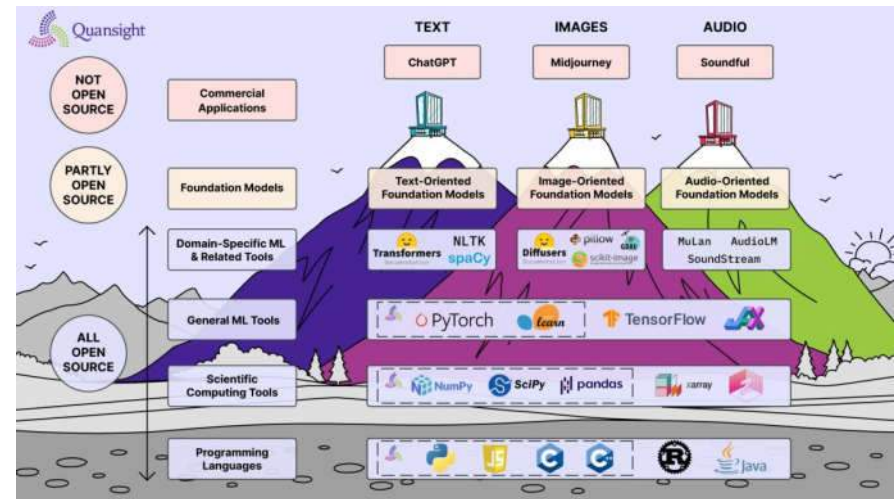
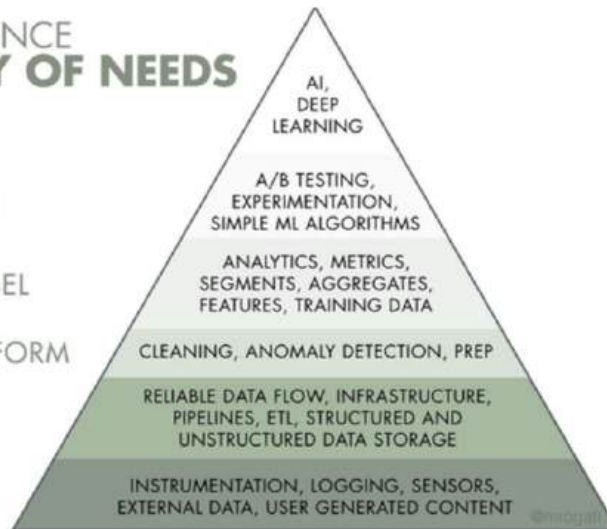
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



Types of Data Professionals

Where are you?

Data Engineer



ML Engineer



Data Scientist



Data Analyst



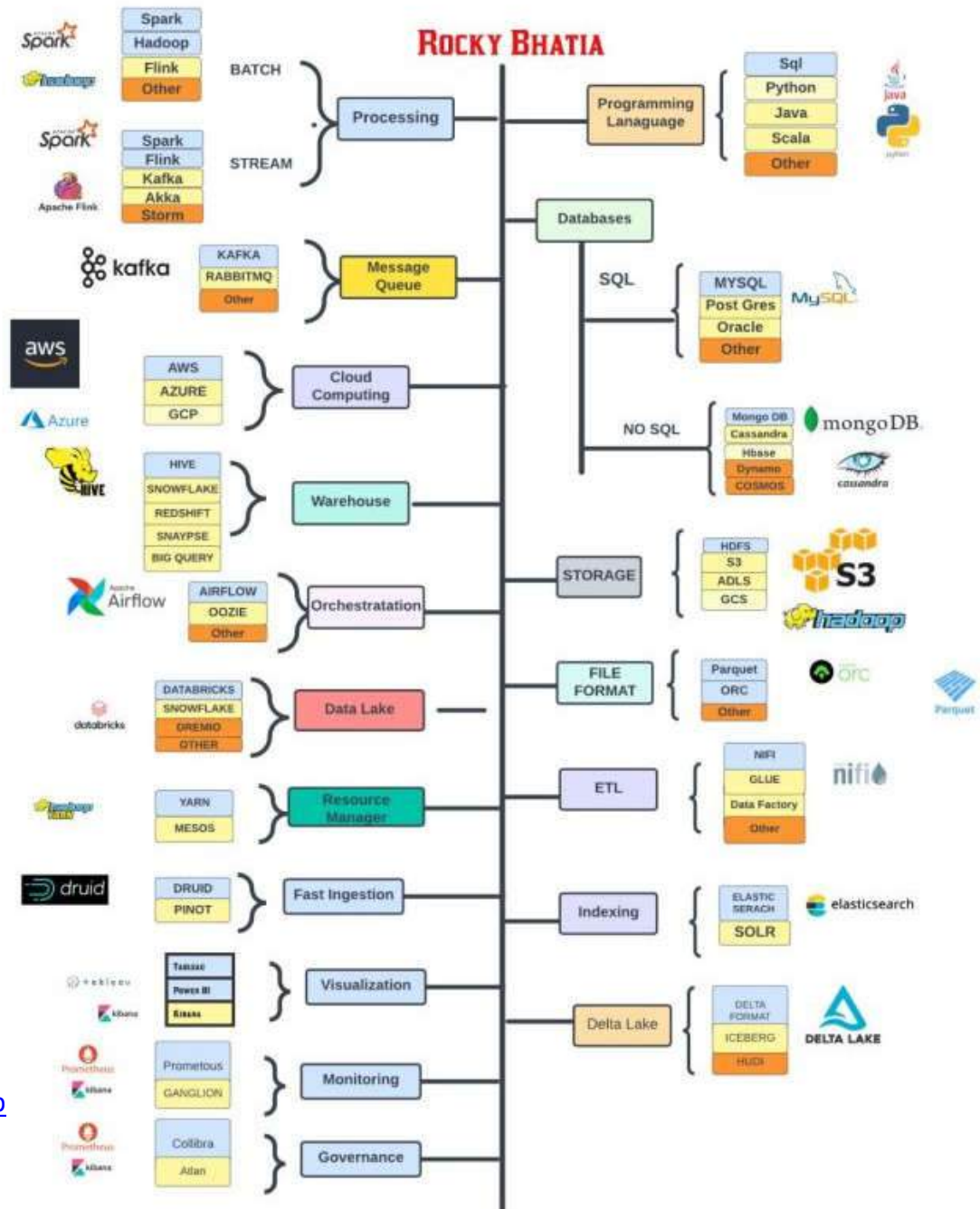
https://x.com/Chandigarh_uni/status/1729537383836667953

Data Eng. Roadmap



DATA ENGINEERING ROADMAP

ROCKY BHATIA

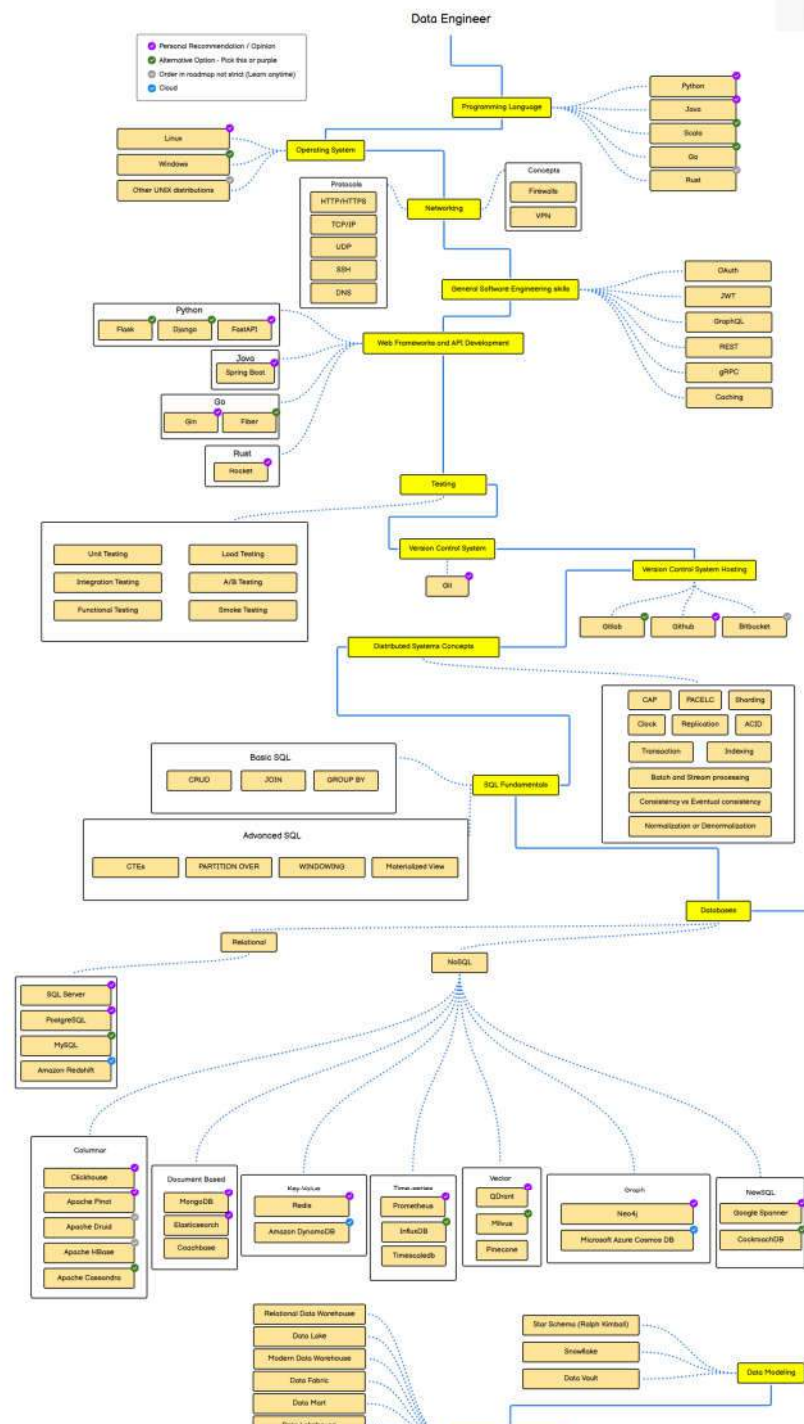
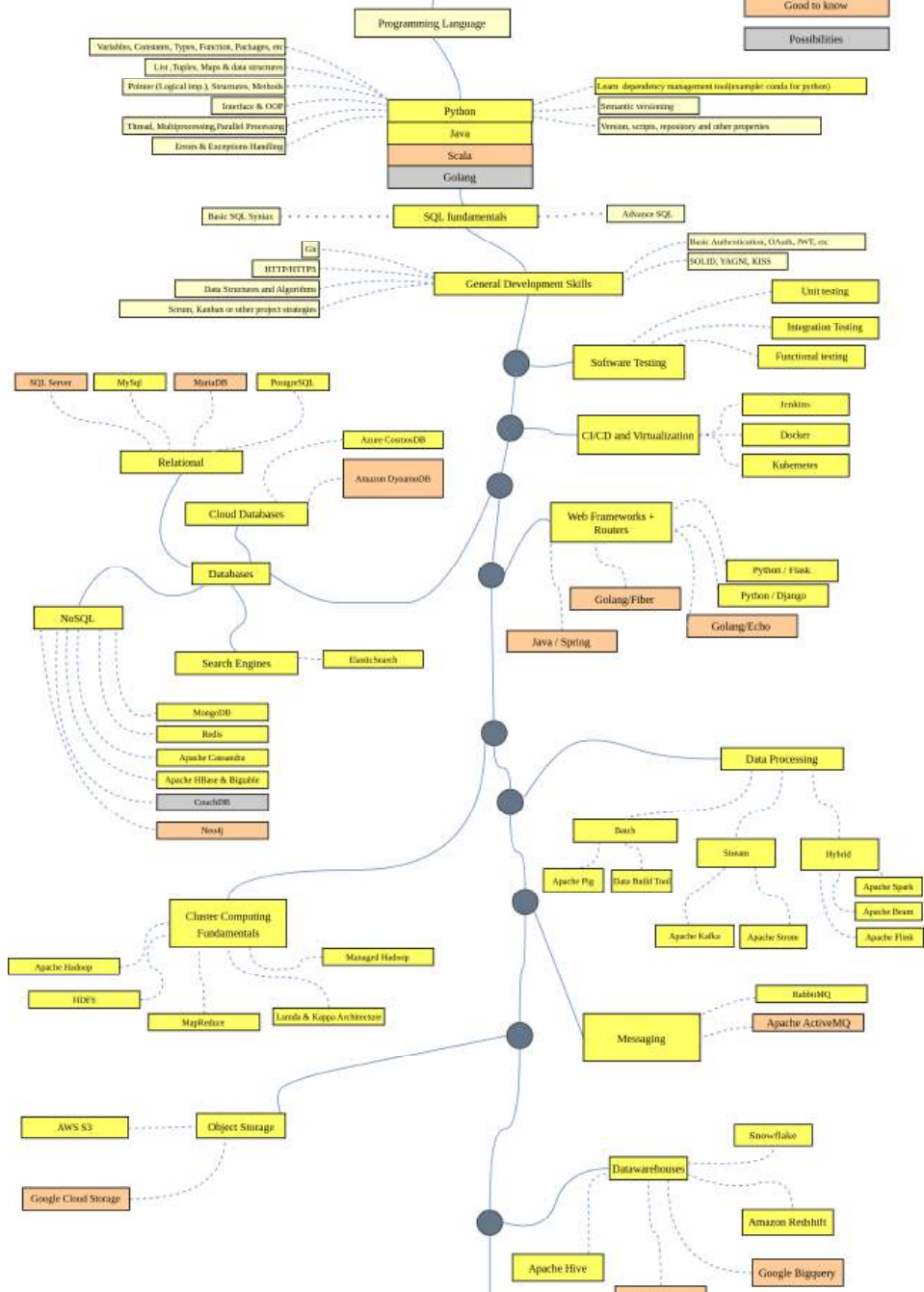


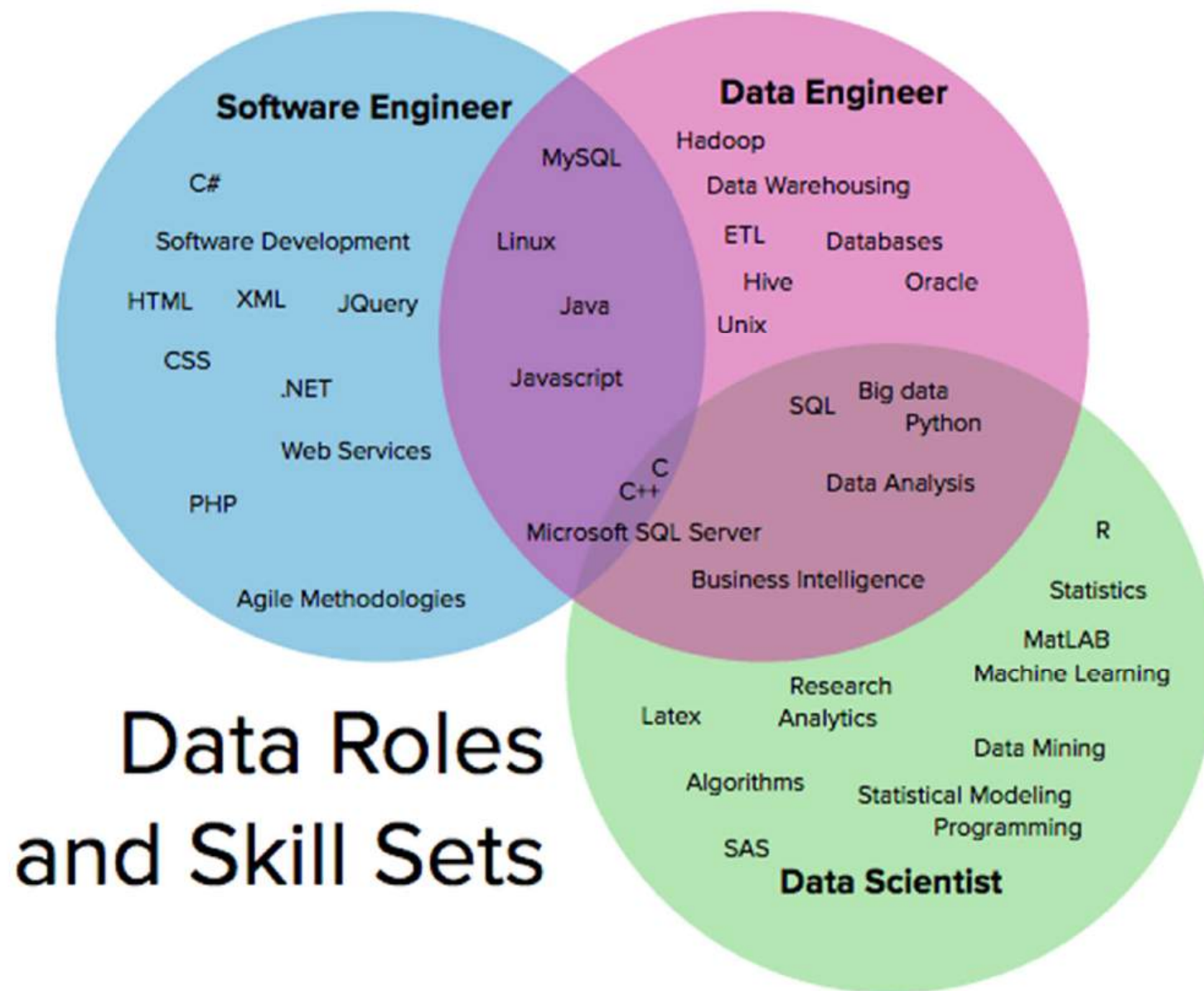
<https://github.com/data-burst/data-engineering-roadmap>
<https://roadmap.sh/>

Data Engineer Roadmap 2022

Legends

- Personal must know
- Good to know
- Possibilities

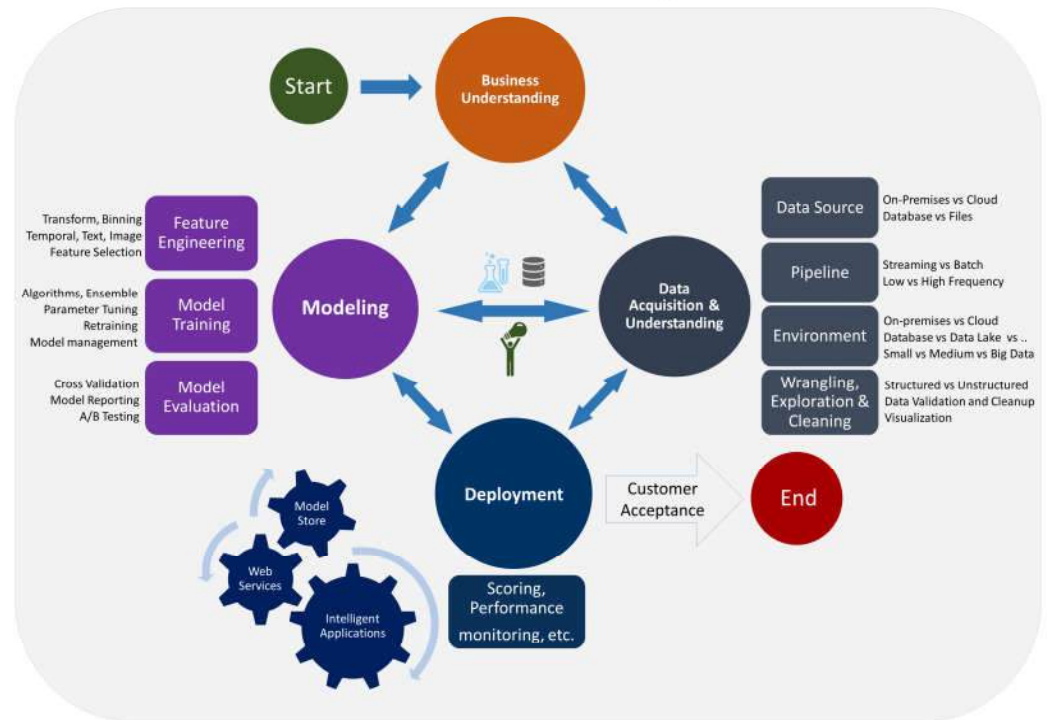




DS needs

- Domain Knowledge
- Data preprocessing (cleaning)
- Data analysis
- Data Analytics
- Missing data handling

Data Science Lifecycle



- LLMs (Chatbots) could help

FIGURE 3.7

Skill importance gap and skill proficiency gap between growing and declining jobs

When growing and declining job roles attach the same level of importance and proficiency to a skill, the index equals one. The bigger the value, the bigger the gap between growing and declining jobs.

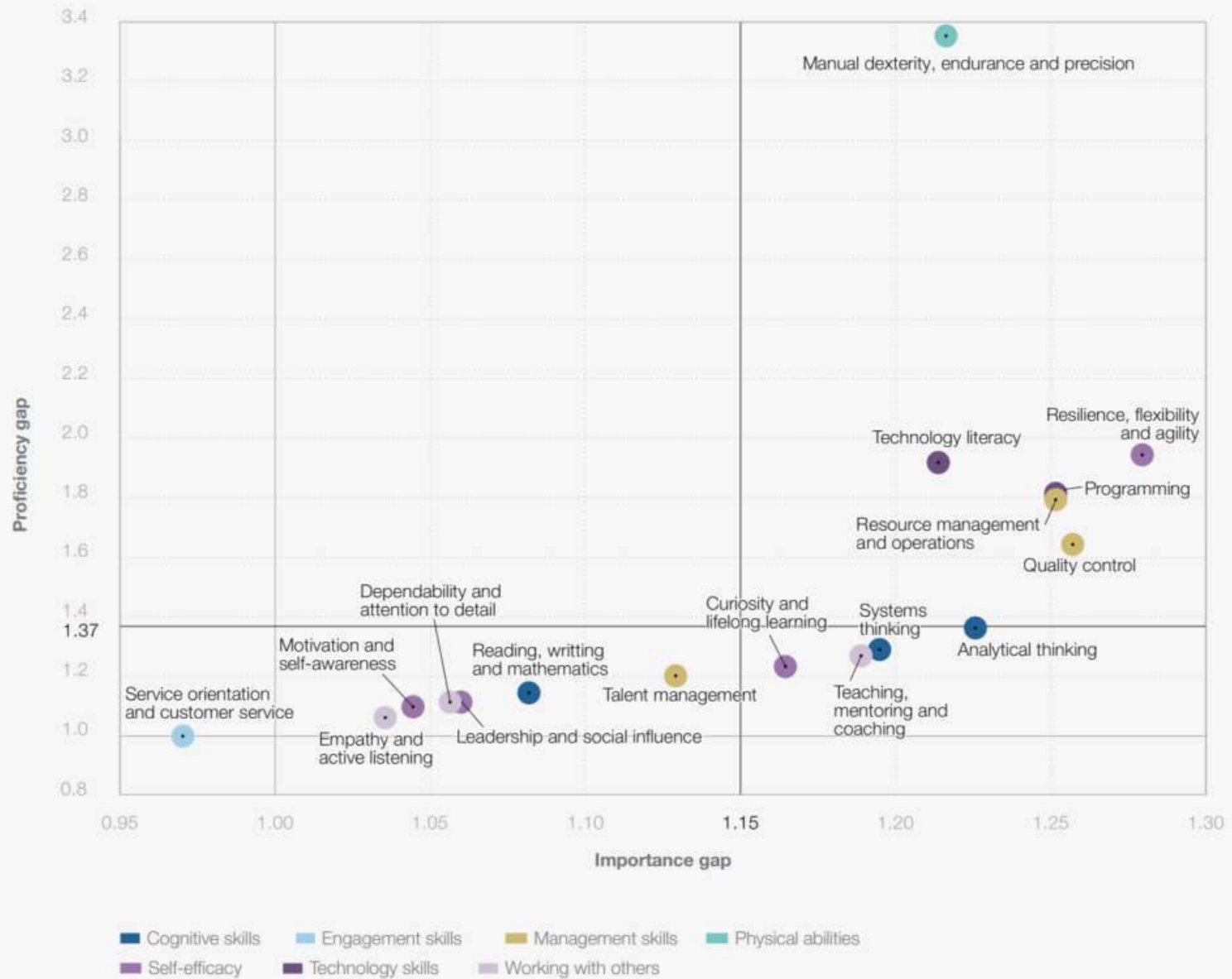
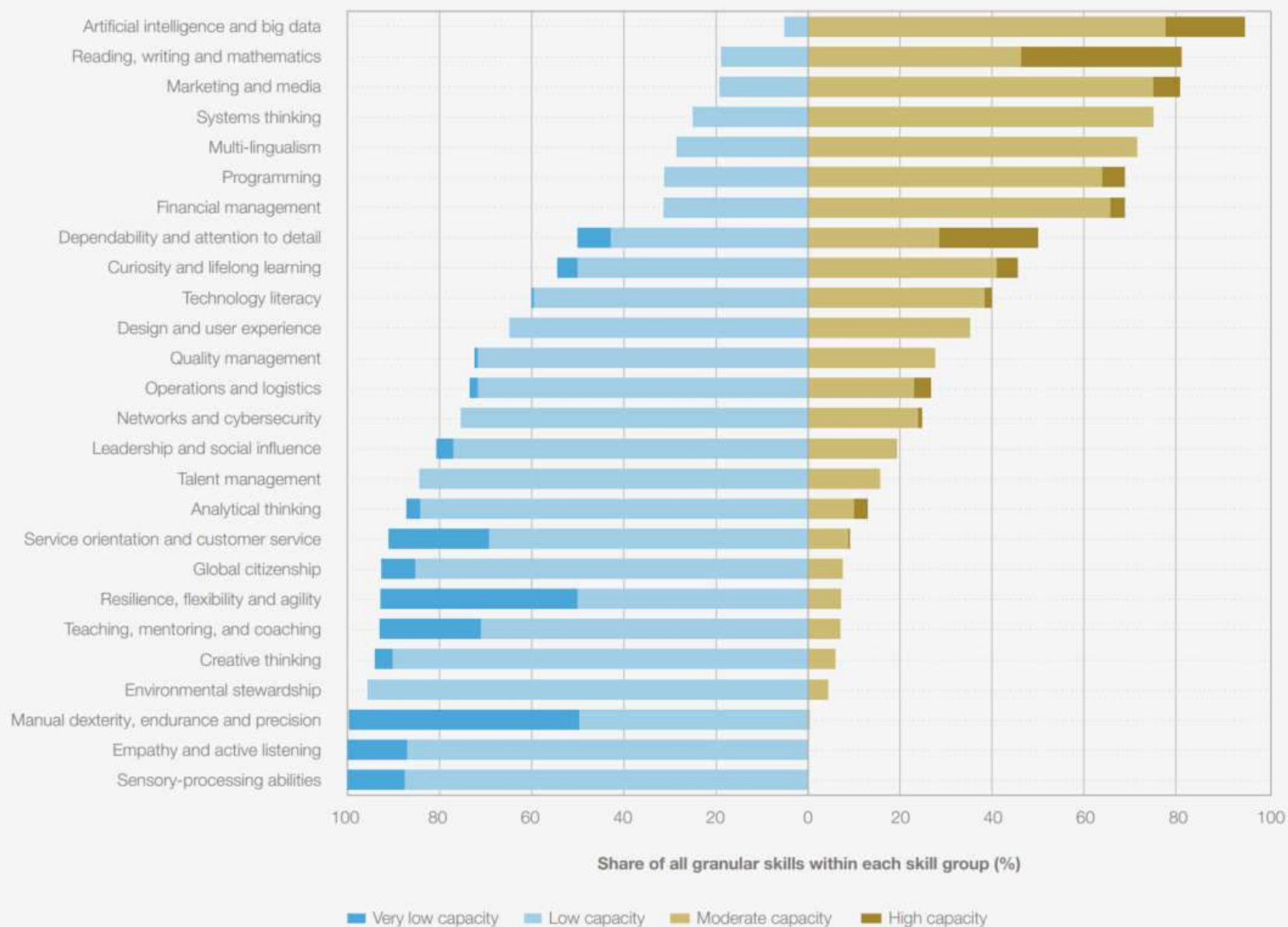


FIGURE B3.1

Current capacity for substitution by Generative AI, by skill group

Capacity of GenAI substituting a human in performing a given skill as a percentage share of all granular skills within each skill group.

Analysis based on GPT-4o, with over 2800 granular skills from the Indeed database as of August 2024.



Skill set

Data Analyst	Data Engineer	Data Scientist
Data Warehousing	Data Warehousing & ETL	Statistical & Analytical skills
Adobe & Google Analytics	Advanced programming knowledge	Data Mining
Programming knowledge	Hadoop-based Analytics	Machine Learning & Deep learning principles
Scripting & Statistical skills	In-depth knowledge of SQL/ database	In-depth programming knowledge (SAS/R/ Python coding)
Reporting & data visualization	Data architecture & pipelining	Hadoop-based analytics
SQL/ database knowledge	Machine learning concept knowledge	Data optimization
Spread-Sheet knowledge	Scripting, reporting & data visualization	Decision making and soft skills

Roles and Responsibilities

Data Analyst	Data Engineer	Data Scientist
Pre-processing and data gathering	Develop, test & maintain architectures	Responsible for developing Operational Models
Emphasis on representing data via reporting and visualization	Understand programming and its complexity	Carry out data analytics and optimization using machine learning & deep learning
Responsible for statistical analysis & data interpretation	Deploy ML & statistical models	Involved in strategic planning for data analytics
Ensures data acquisition & maintenance	Building pipelines for various ETL operations	Integrate data & perform ad-hoc analysis
Optimize Statistical Efficiency & Quality	Ensures data accuracy and flexibility	Fill in the gap between the stakeholders and customer

سوال و جواب

AI effect

