The Future of Data Engineering

in a Post-AI world

Michelle Winters
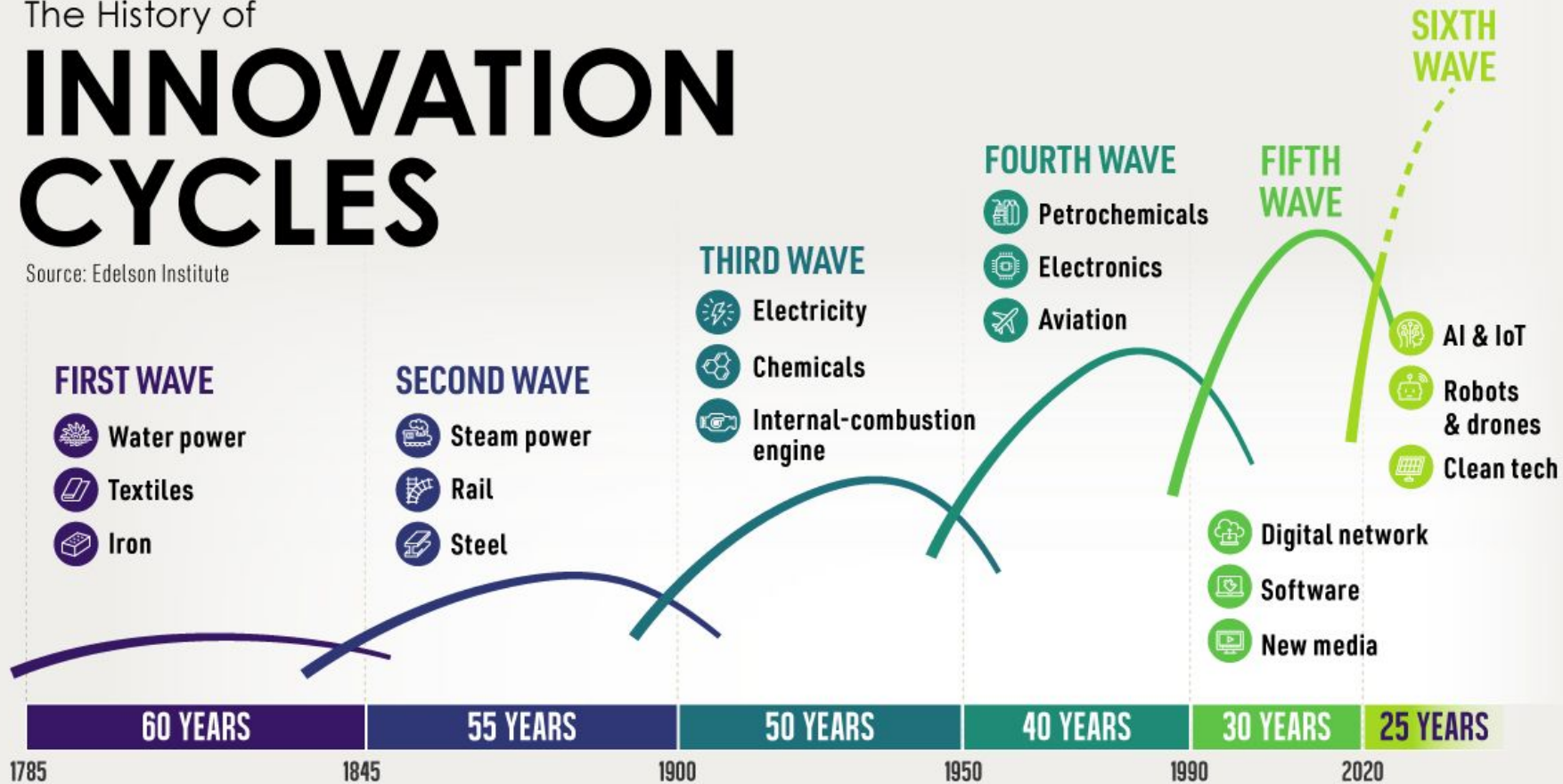
Distinguished Architect, eBay

AI is here.

DALL·E 3

The History of
INNOVATION CYCLES

Source: Edelson Institute

FIRST WAVE
- Water power
- Textiles
- Iron

SECOND WAVE
- Steam power
- Rail
- Steel

THIRD WAVE
- Electricity
- Chemicals
- Internal-combustion engine

FOURTH WAVE
- Petrochemicals
- Electronics
- Aviation

FIFTH WAVE
- Digital network
- Software
- New media

SIXTH WAVE
- AI & IoT
- Robots & drones
- Clean tech

60 YEARS	55 YEARS	50 YEARS	40 YEARS	30 YEARS	25 YEARS

1785   1845   1900   1950   1990   2020

Source: VISUAL CAPITALIST

D Data Council

Who am I?

GoDaddy

NETFLIX

noteable

ebay

A New Frontier of Analytics Maturity

Cognitive Analytics

1 AI: "Let me help you with that."

Prescriptive Analytics

2 ML: "What should we do next?"

Predictive Analytics

3 Data Science: "What's likely to happen?"

Diagnostic Analytics

4 Analytics: "Why did it happen?"

Value

Descriptive Analytics

5 Reporting: "What happened?"

→ "What's happening **now**?"
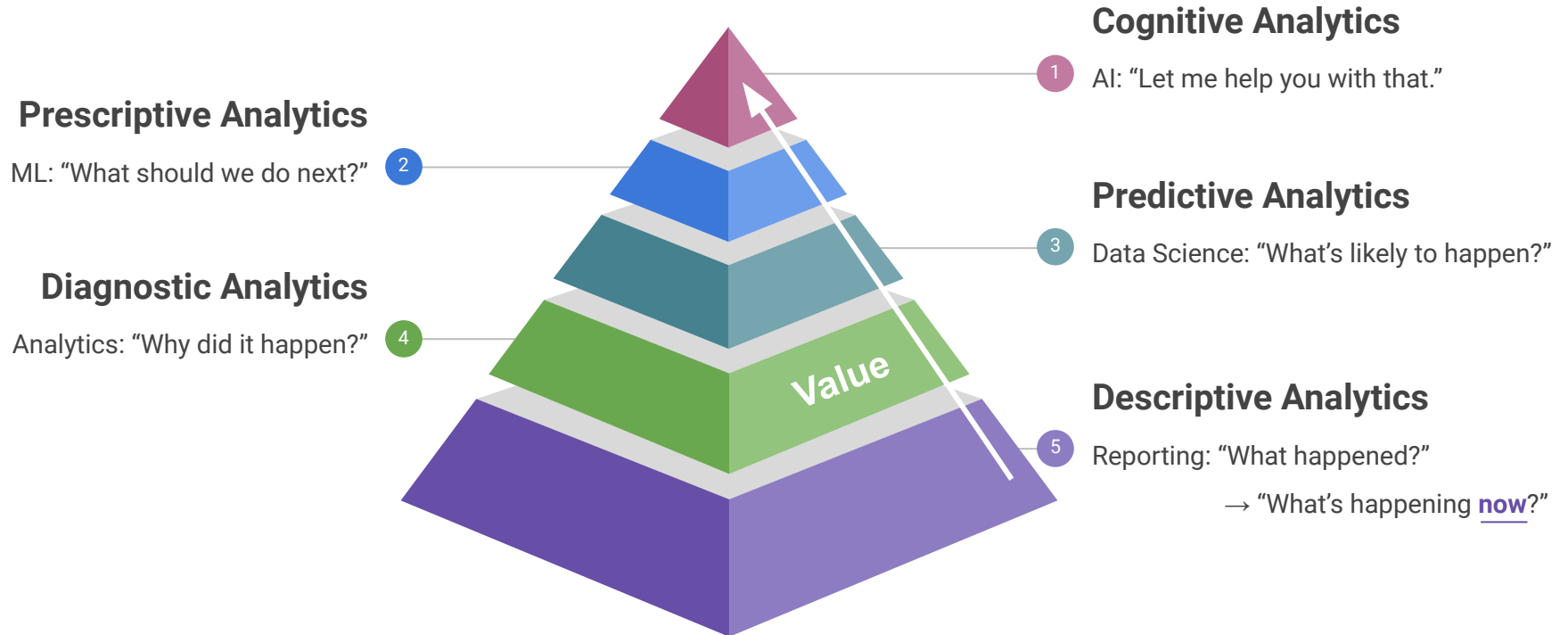
"New" Data Technologies Powering AI

LLM

fine-tuning
enrichment

retrieval
augmentation

real-time

embeddings

V1 = [-0.01152326, -0.02584043, ...]
V2 = [-0.00072085, -0.01380217, ...]

Knowledge
Graphs

Vector
Databases

Input Topic

Long-Term
Storage

Stream
Processing

Query

Kappa
Architecture

Data Council

MWinters 2024
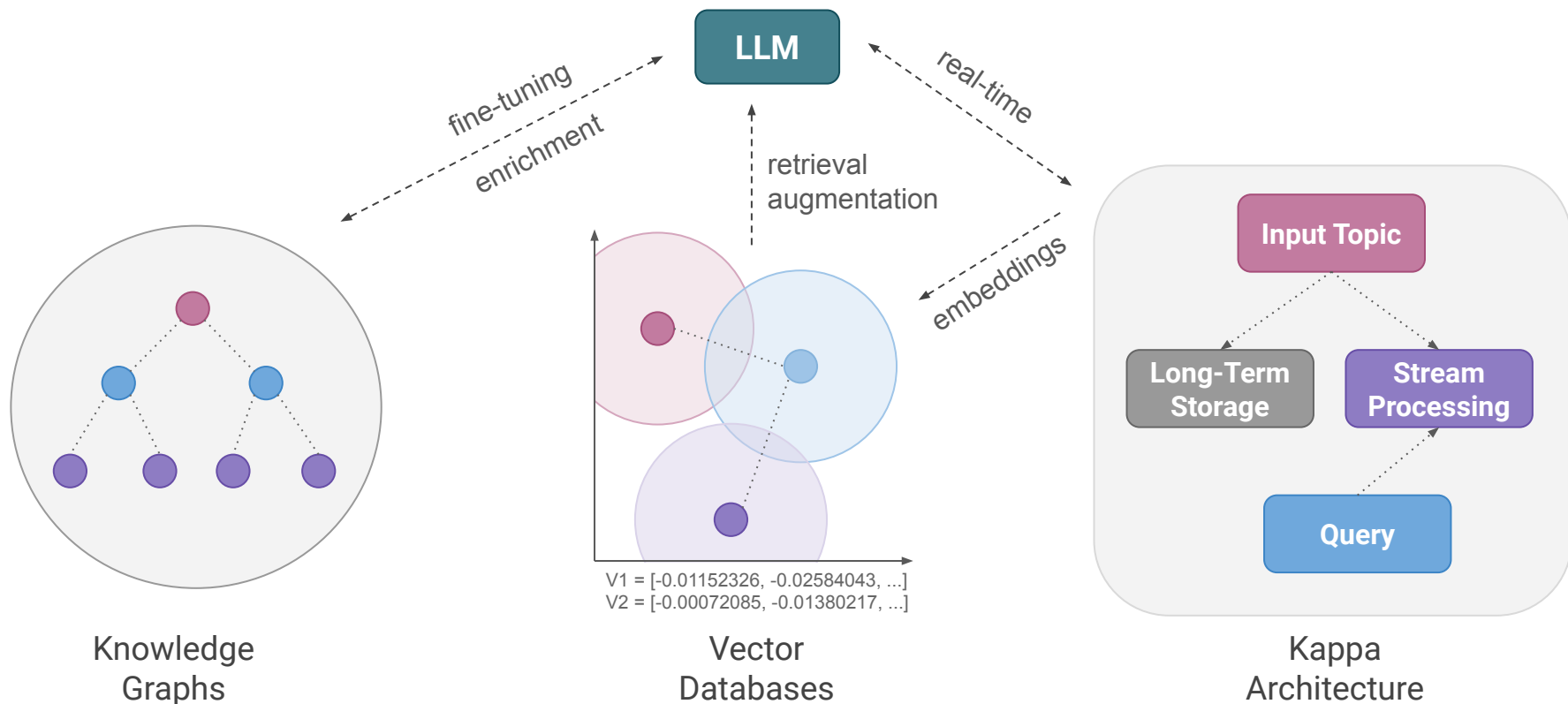
Relational is dead. Long live relational.

418 systems in ranking, March 2024

	Rank		DBMS	Database Model		Score	
Mar 2024	Feb 2024	Mar 2023			Mar 2024	Feb 2024	Mar 2023
1.	1.	1.	Oracle ➕	Relational, Multi-model ℹ️	1221.06	-20.39	-40.23
2.	2.	2.	MySQL ➕	Relational, Multi-model ℹ️	1101.50	-5.17	-81.29
3.	3.	3.	Microsoft SQL Server ➕	Relational, Multi-model ℹ️	845.81	-7.76	-76.20
4.	4.	4.	PostgreSQL ➕	Relational, Multi-model ℹ️			08
5.	5.	5.	MongoDB ➕	Document, Multi-model ℹ️			25
6.	6.	6.	Redis ➕	Key-value, Multi-model ℹ️			45
7.	7.	⬆ 8.	Elasticsearch	Search engine, Multi-model			28
8.	8.	⬇ 7.	IBM Db2	Relational, Multi-model ℹ️			17
9.	9.	⬆ 11.	Snowflake ➕	Relational	125.38	-2.07	+10.98
10.	10.	⬇ 9.	SQLite ➕	Relational	118.16	+0.88	-15.66
11.	11.	⬇ 10.	Microsoft Access	Relational	107.93	-5.24	-24.13
12.	12.	12.	Cassandra ➕	Wide column, Multi-model ℹ️	104.59	-4.69	-9.20
13.	13.	13.	MariaDB ➕	Relational, Multi-model ℹ️	95.03	-2.20	-1.81
14.	14.	14.	Splunk	Search engine	89.68	-1.97	+1.71
15.	⬆ 16.	⬆ 16.	Microsoft Azure SQL Database	Relational, Multi-model ℹ️	78.51	-1.06	+1.06
16.	⬇ 15.	⬇ 15.	Amazon DynamoDB ➕	Multi-model ℹ️	77.72	-5.18	-3.05
17.	17.	⬆ 19.	Databricks ➕	Multi-model ℹ️	74.34	-2.57	+13.48
18.	18.	⬇ 17.	Hive	Relational	64.82	-0.99	-6.09
19.	19.	⬆ 21.	Google BigQuery ➕	Relational	62.67	-0.96	+9.23
20.	20.	⬇ 18.	Teradata	Relational, Multi-model ℹ️	48.95	-2.29	-14.79

Relational DBMS
Document store
Graph DBMS
Spatial DBMS
Vector DBMS

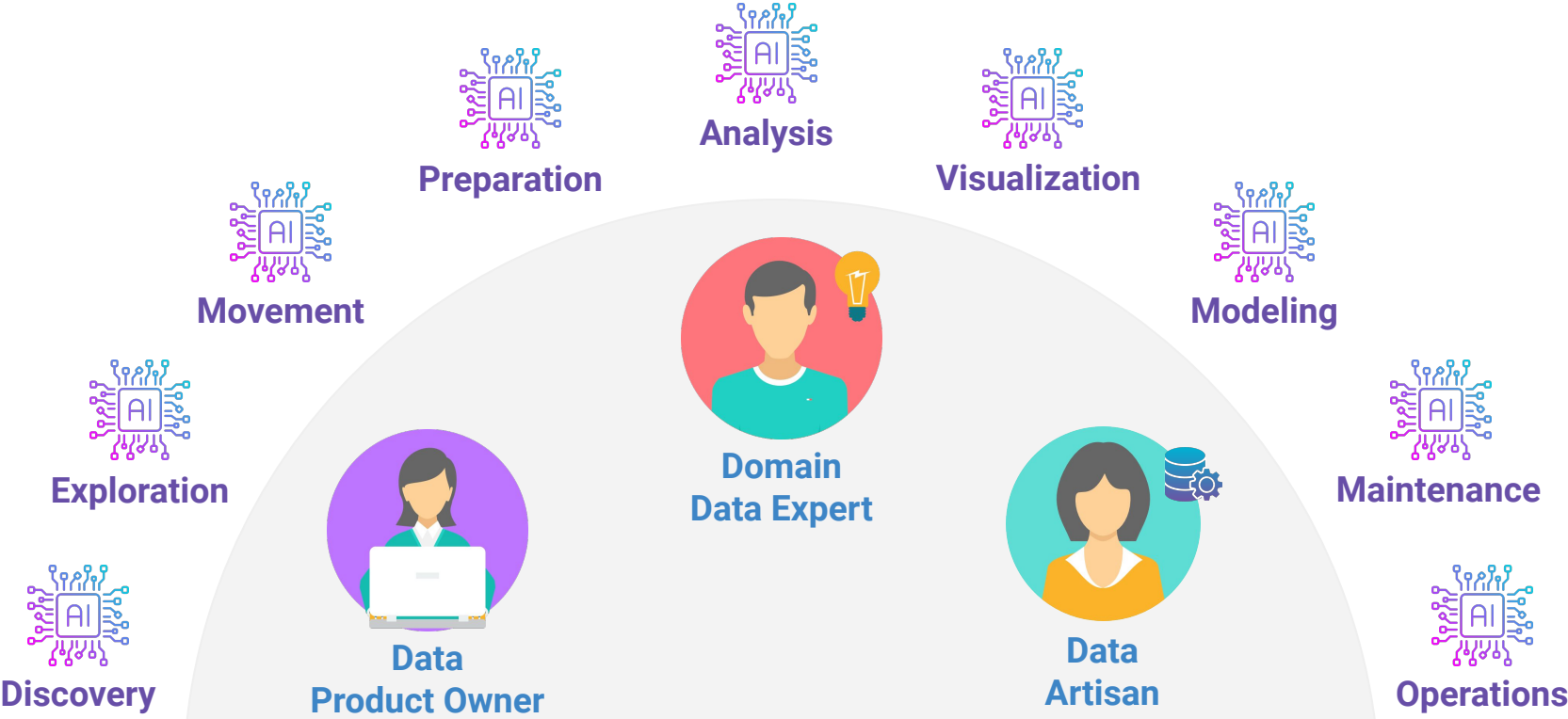Source:

DB-ENGINES

The Modern Data Stack, Circa 2024

DALL·E 3

DALL·E 3
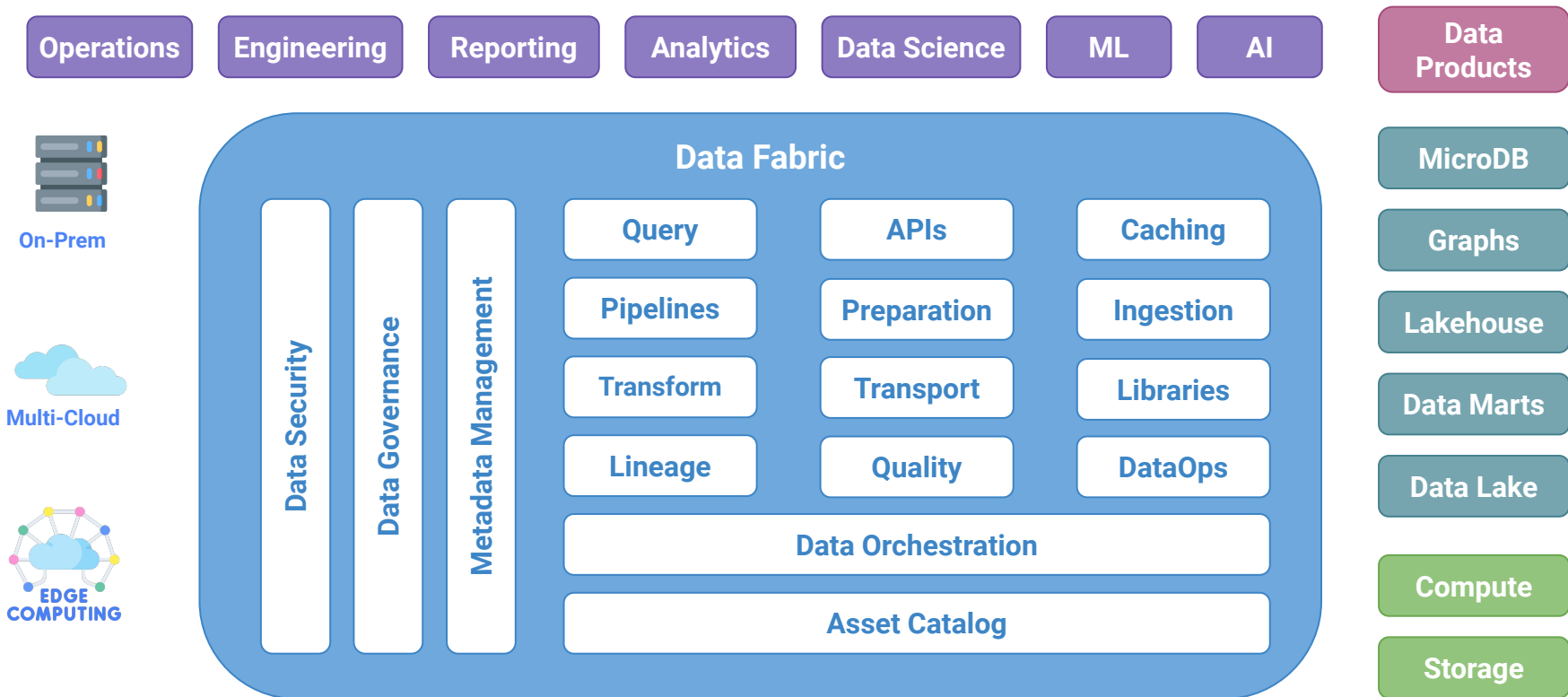
A NETFLIX ORIGINAL SERIES

Data Council

[Quantum + Graphene + AI] = 6th Industrial Revolution
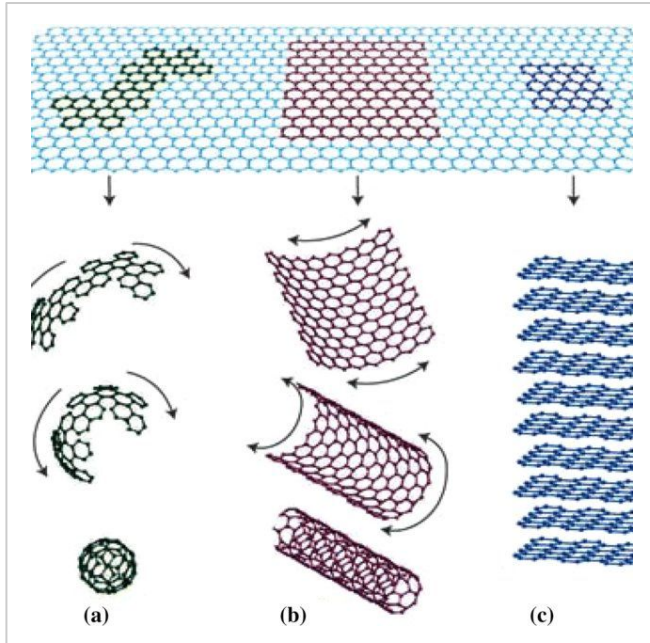

Image Source: Amit Kumar, Kamal Sharma, Amit Rai Dixit
Journal of Materials Science, Apr 2019

Quantum computing is **158 million times** faster than today's supercomputers.

Graphene is the **lightest**, **strongest, most energy efficient** material ever discovered. It's the material of the future.

AI provides the **intelligence** to connect our **digital fabric** to our physical plane in a meaningful way.

The best is yet to come.

AI can help us build
a better tomorrow.

Data Council

DALL·E 3

Thank you
for listening.

———

Michelle Winters

linkedin.com/in/mufford