

DATA SCIENCE ON TOP OF AWS

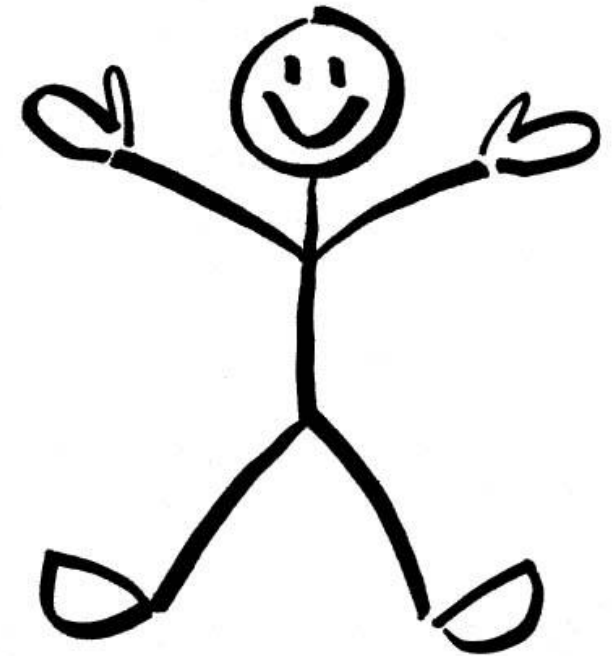


Carlos Badillo  
Gabriela Martinez

GitHub repo: <https://github.com/mgmartinezl/Research-topic>

"THE SEXIEST JOB OF THE 21<sup>ST</sup> CENTURY"

A DATA SCIENTIST KNOWS MORE  
STATISTICS THAN A COMPUTER  
SCIENTIST, BUT MORE ABOUT  
COMPUTER HACKING THAN A  
STATISTICIAN

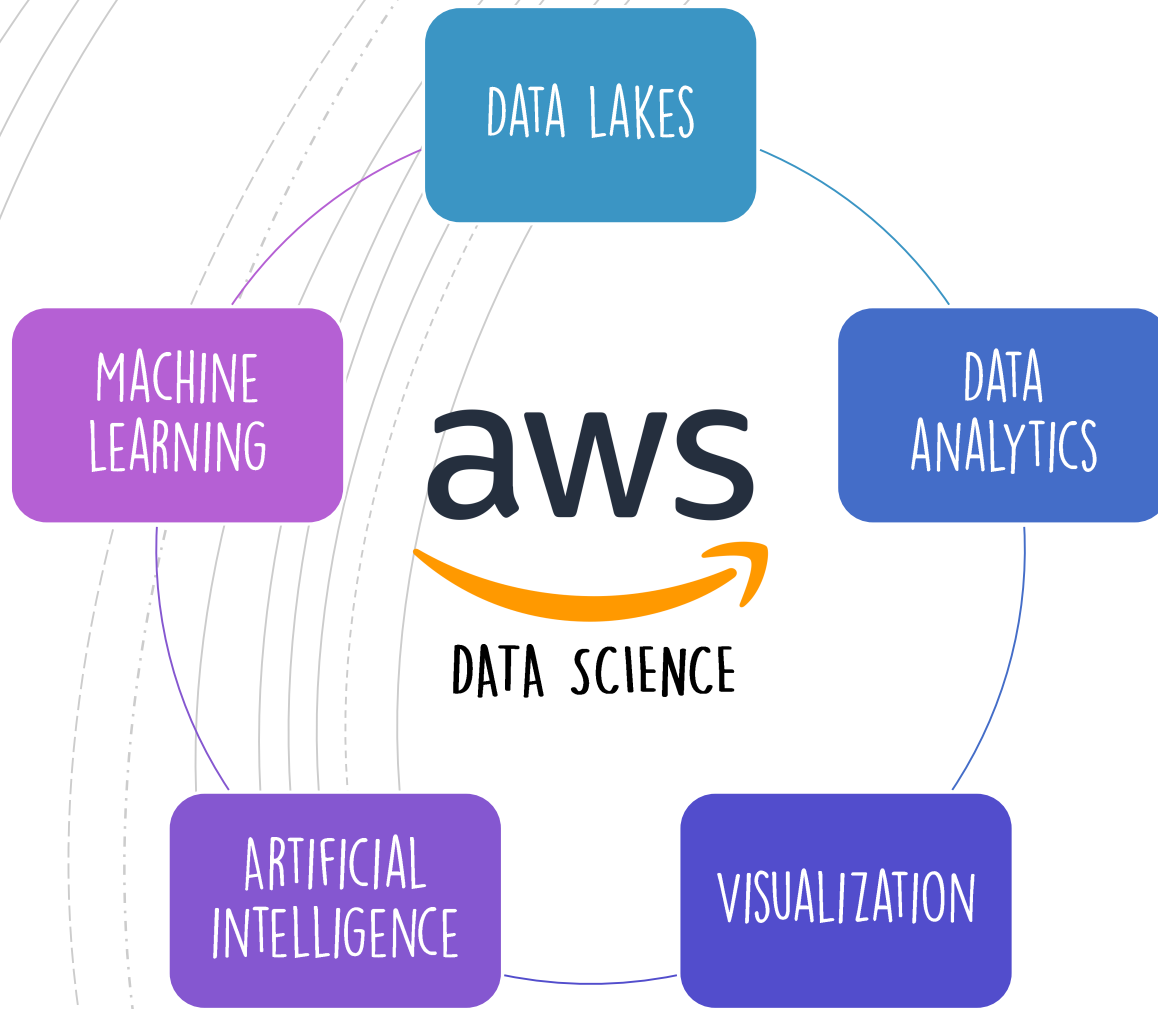


DATA SCIENTIST

WHY IS THIS MIX IMPORTANT?

# DATA SCIENCE ENABLES BETTER DECISION-MAKING

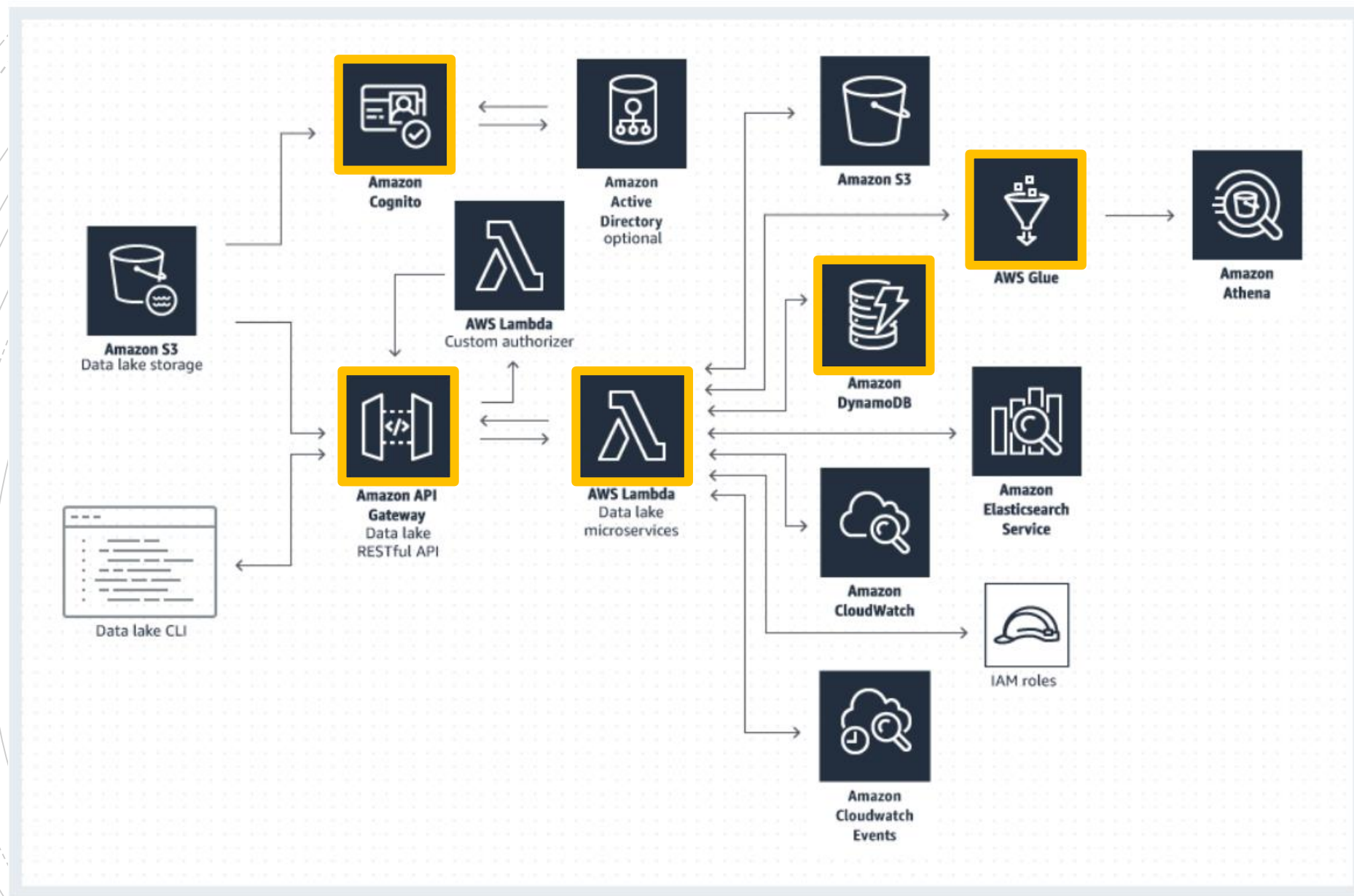


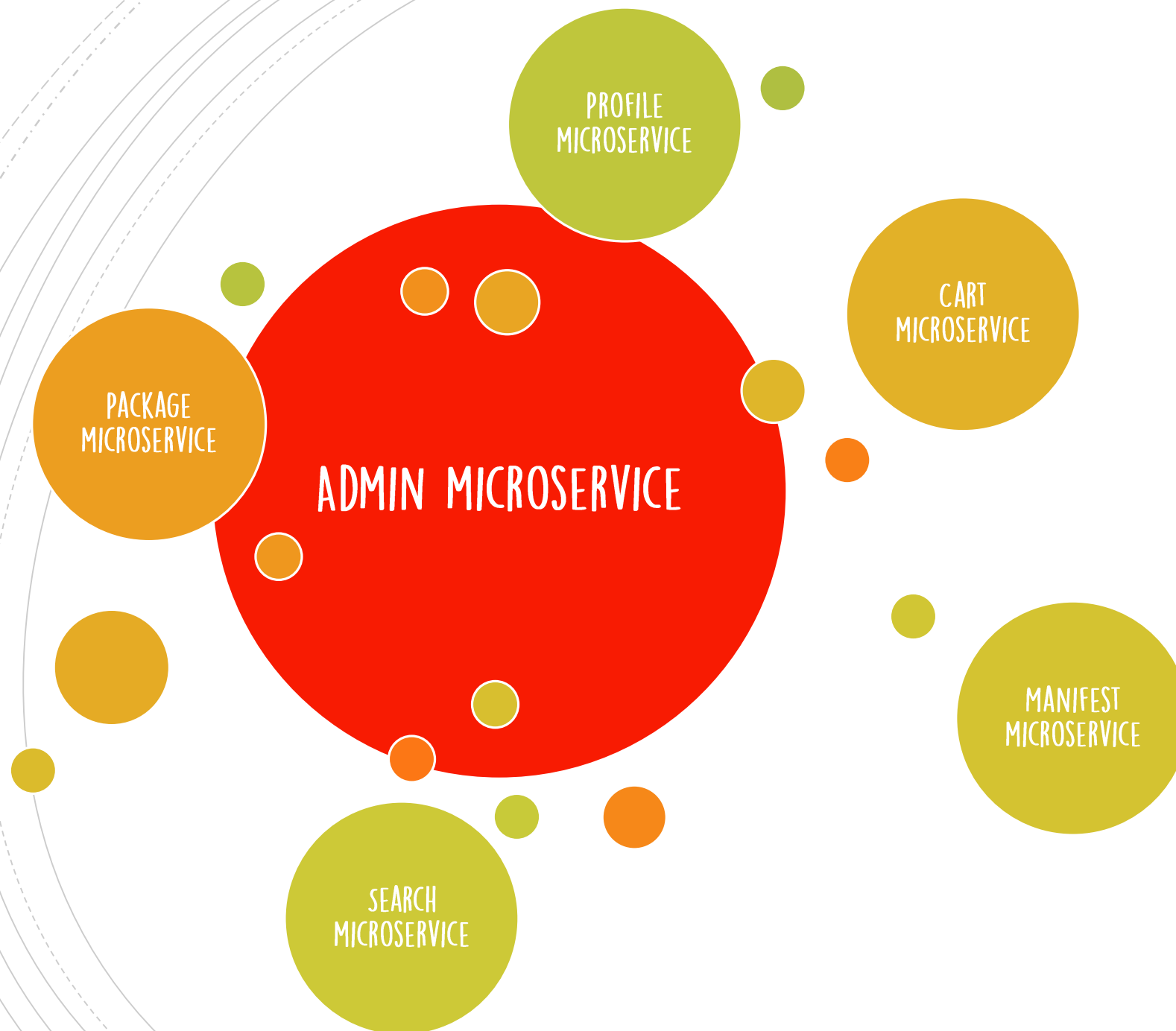


"DATA SCIENCE IS THE CIVIL  
ENGINEERING OF DATA.

ITS ACOLYTES POSSESS A  
PRACTICAL KNOWLEDGE OF  
TOOLS AND MATERIALS, COUPLED  
WITH A THEORETICAL  
UNDERSTANDING OF WHAT'S  
POSSIBLE" MIKE DRISCOLL'S

# DATA LAKES ON TOP OF AWS





# ANALYTICS ON TOP OF AWS



INTERACTIVE ANALYTICS

- AMAZON ATHENA  
QUERY YOUR DATA LAKE WITH STANDARD SQL



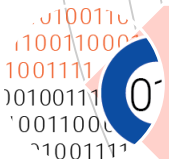
BIG DATA PROCESSING



DATA WAREHOUSING



REAL-TIME ANALYTICS



OPERATIONAL ANALYTICS



# ANALYTICS ON TOP OF AWS



INTERACTIVE ANALYTICS



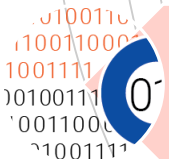
BIG DATA PROCESSING



DATA WAREHOUSING



REAL-TIME ANALYTICS



OPERATIONAL ANALYTICS

- AMAZON EMR

EASILY RUN AND SCALE APACHE SPARK, HADOOP, HBASE, PRESTO, HIVE, AND OTHER BIG DATA FRAMEWORKS



# ANALYTICS ON TOP OF AWS



INTERACTIVE ANALYTICS



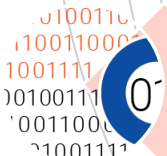
BIG DATA PROCESSING



DATA WAREHOUSING



REAL-TIME ANALYTICS



OPERATIONAL ANALYTICS

- AMAZON REDSHIFT

SCALABLE DATA WAREHOUSE WITH  
TEN TIMES FASTER PERFORMANCE

# ANALYTICS ON TOP OF AWS



INTERACTIVE ANALYTICS



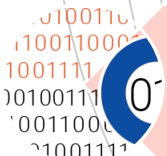
BIG DATA PROCESSING



DATA WAREHOUSING



REAL-TIME ANALYTICS



OPERATIONAL ANALYTICS

- AMAZON KINESIS  
COLLECT, PROCESS, AND ANALYZE VIDEO  
AND DATA STREAMS IN REAL TIME

# ANALYTICS ON TOP OF AWS



INTERACTIVE ANALYTICS



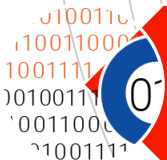
BIG DATA PROCESSING



DATA WAREHOUSING



REAL-TIME ANALYTICS



OPERATIONAL ANALYTICS

- AMAZON ELASTICSEARCH SERVICE

# AMAZON QUICKSIGHT



## Collect and load data

Clickstreams, sales orders, IoT, financial data, and more



## Data sources

Seamlessly connect to your data wherever it lives - in the cloud, in 3rd party applications, or on-premises



## Amazon QuickSight

First BI service with Pay-per-Session pricing



Interactive dashboards



Email reports



Embedded analytics



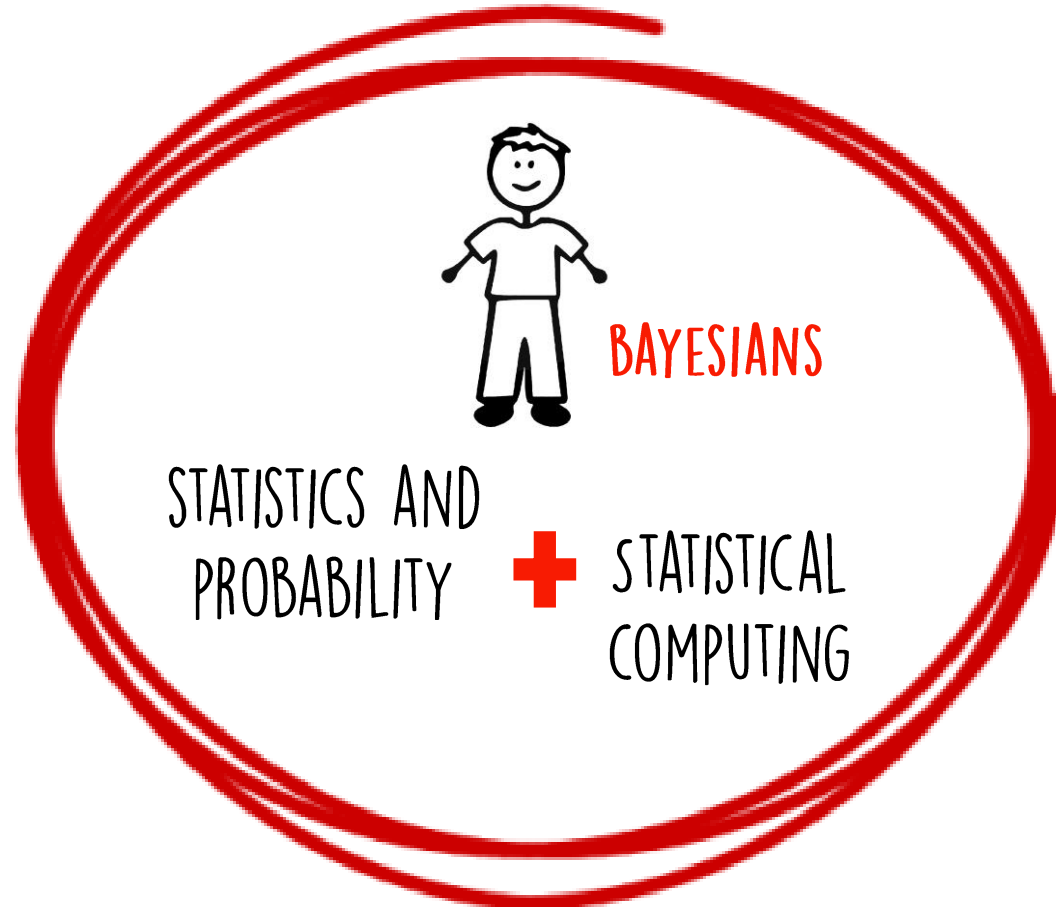
Based on [\[3\]](#)

# VISUALIZATION ON TOP OF AWS

# ARTIFICIAL INTELLIGENCE

"...IS THE FIELD OF COMPUTER SCIENCE  
DEDICATED TO SOLVING COGNITIVE  
PROBLEMS COMMONLY ASSOCIATED  
WITH HUMAN INTELLIGENCE, SUCH AS  
LEARNING, PROBLEM SOLVING, AND  
PATTERN RECOGNITION"

# MACHINE LEARNING



- ANOMALY DETECTION
- FRAUD DETECTION
- CUSTOMER CHURN
- CONTENT PERSONALIZATION

MACHINE LEARNING

# MACHINE LEARNING ON TOP OF AWS

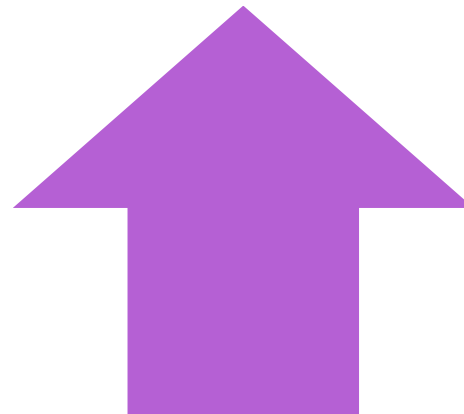
## AMAZON SAGEMAKER



- \*MULTIFRAMEWORK
- \*PRE-BUILT AND TAILORED MODELS
- \*REDUCES TIME-TO-MARKET
- \*INTUITIVE VISUALIZATION



- \*INTERFACE
- \*LACKS VAST DOCUMENTATION
- \*CAN BE EXPENSIVE DUE TO GPU OPTIMIZATION
- \*NOT POSSIBLE TO SCHEDULE TRAINING JOBS





# DEEP LEARNING

DEEP  
LEARNING



CONNECTIONISTS

NEUROSCIENCE



NETWORK  
COMPUTATION



DOES NOT EXPLAIN RELATIONSHIPS  
BETWEEN VARIABLES, AS MACHINE  
LEARNING DOES.



IT FINDS UNKNOWN PATTERNS BY  
MEANS OF THE LAYERS OF **NON-  
LINEAR ALGORITHMS**

# DEEP LEARNING ON TOP OF AWS: **AMAZON REKOGNITION**



Object, scene, and activity detection



Facial recognition

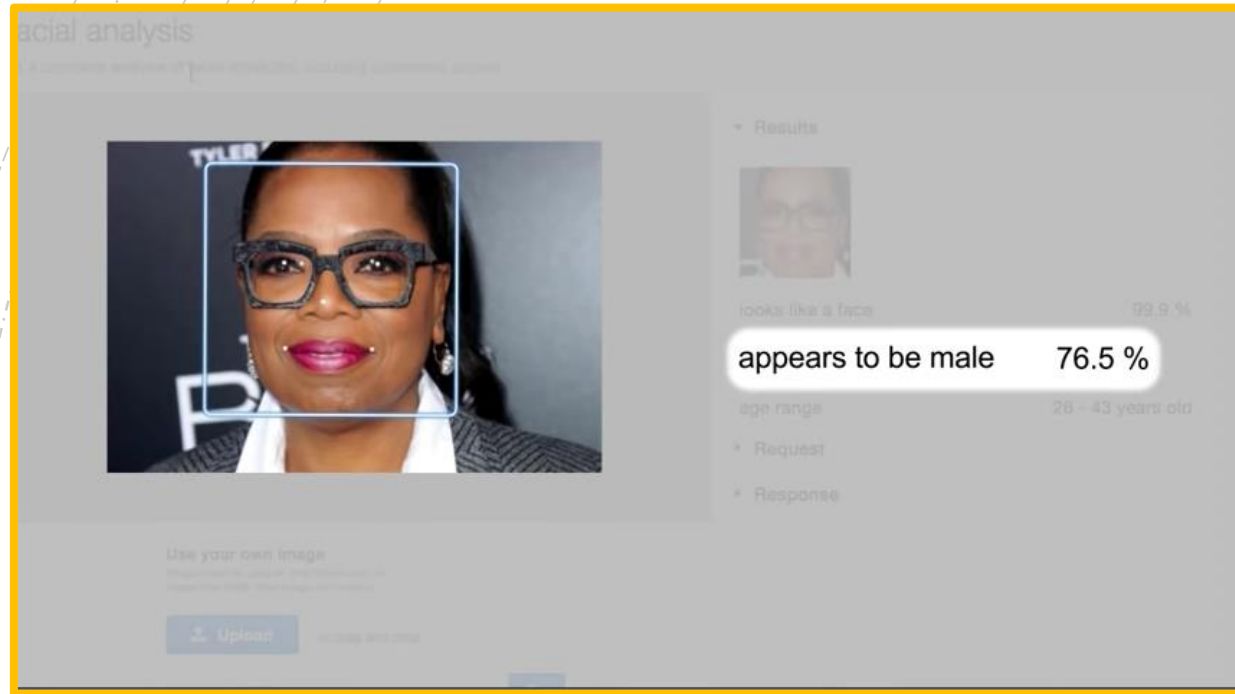


Facial analysis



Pathing

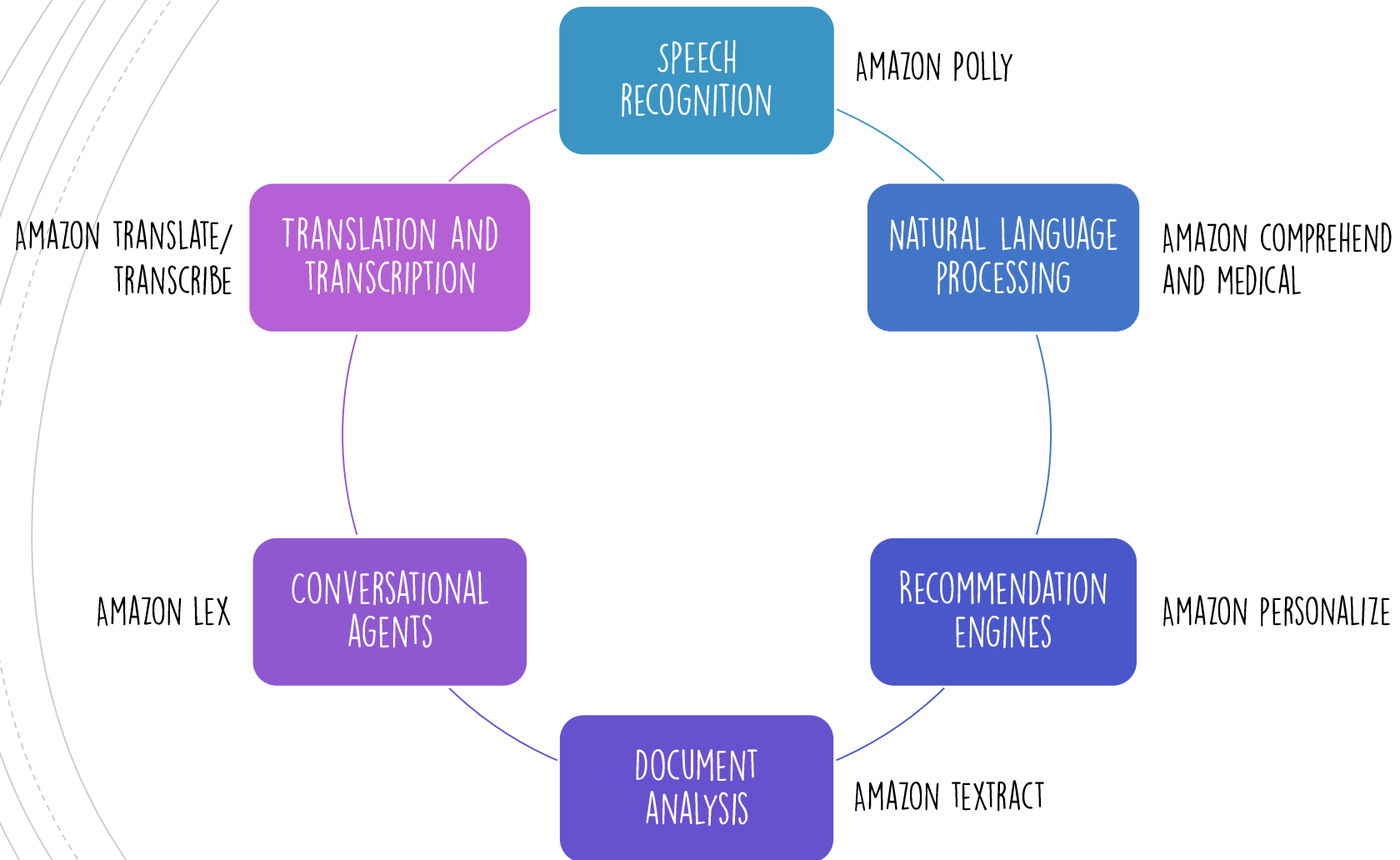
# DEEP LEARNING ON TOP OF AWS: AMAZON REKOGNITION



OPRAH WINFREY

ACCORDING TO MIT RESEARCHERS, AMAZON REKOGNITION **FAILS UP TO 20% MORE** WHEN DETECTING BLACK AND ASIAN FACES (EVEN FOR CELEBRITIES!!!).

# OTHER DEEP LEARNING APPLICATIONS







# BUSINESS CASE

## FORMULA 1

- 120 SENSORS ON EACH CAR
- 3 GB OF DATA
- 1500 OBSERVATIONS EVERY SECOND

## AMAZON STREAMING KINESIS

- PERFORMANCE STATISTICS
- REAL-TIME RACE PREDICTIONS
- PROVIDE FANS WITH INSIGHTS