

AI- and Data-Driven Support For FINTECH

Beng Chin OOI

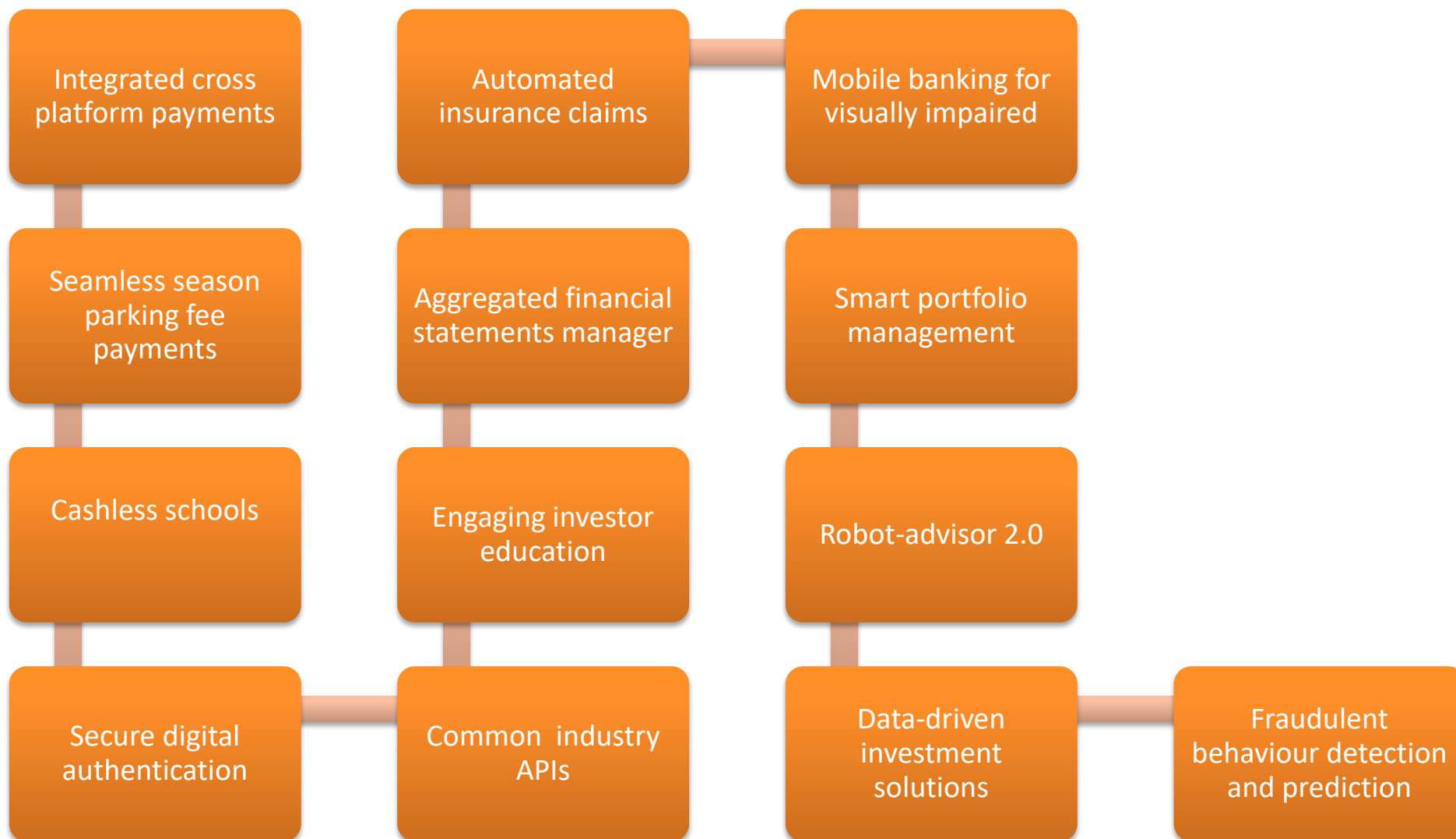
www.comp.nus.edu.sg/~ooibc



Contents

- **AI and QoS**
- **GIT for Data: Facilitating AI Services without Data Sharing**
 - **Our Software Stack**
- **Financial News Analytics**
 - **An Example on using AI/ML**

MAS Fintech Challenges



The Reality of Exploiting AI

The actual implementation of the ML algorithm is usually **less than 5%** lines of code in a real, non-trivial application

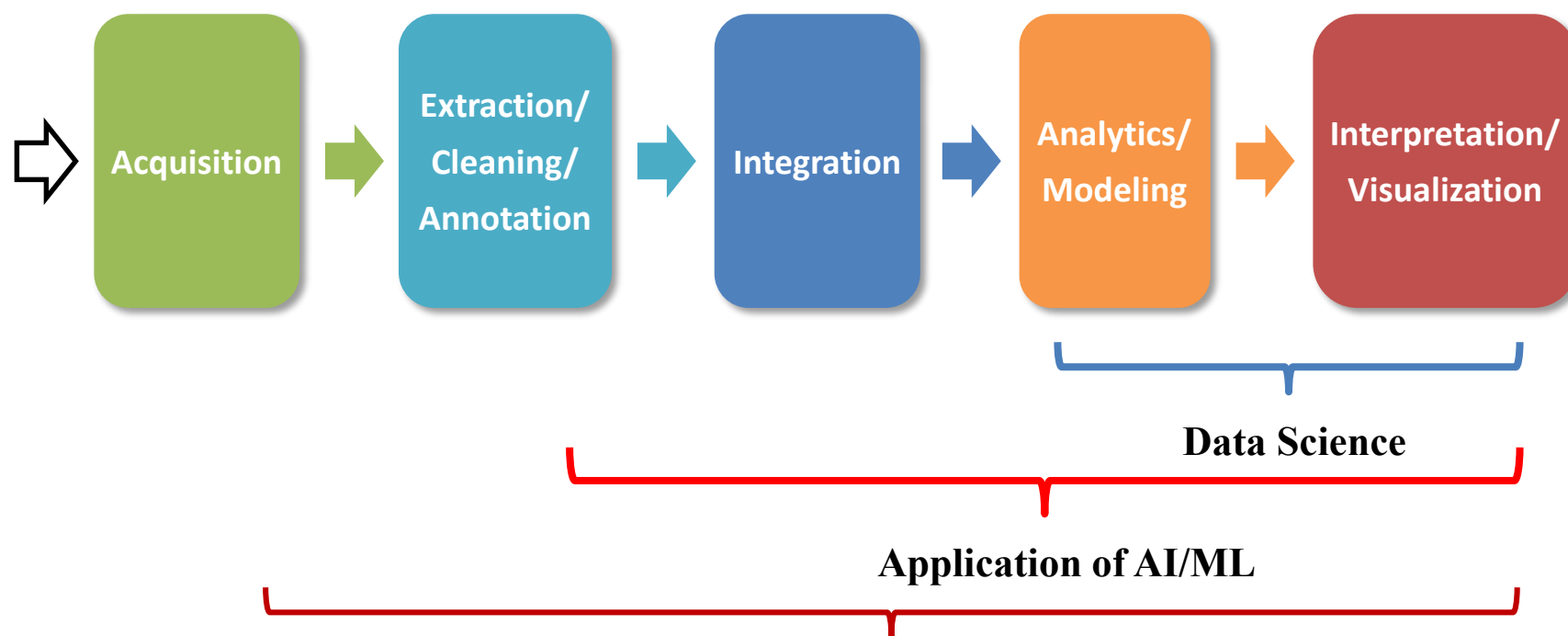
The main effort (i.e. those 95% LOC) is spent on:

- Data cleaning & annotation
- Data extraction, transformation, loading
- Data integration & pruning
- Parameter tuning
- Model training & deployment
-

**These are what we have
been doing!**

This blurs the line between DB and “non-DB” processing, and calls for better integration

The BIG Data Analytics Pipeline*



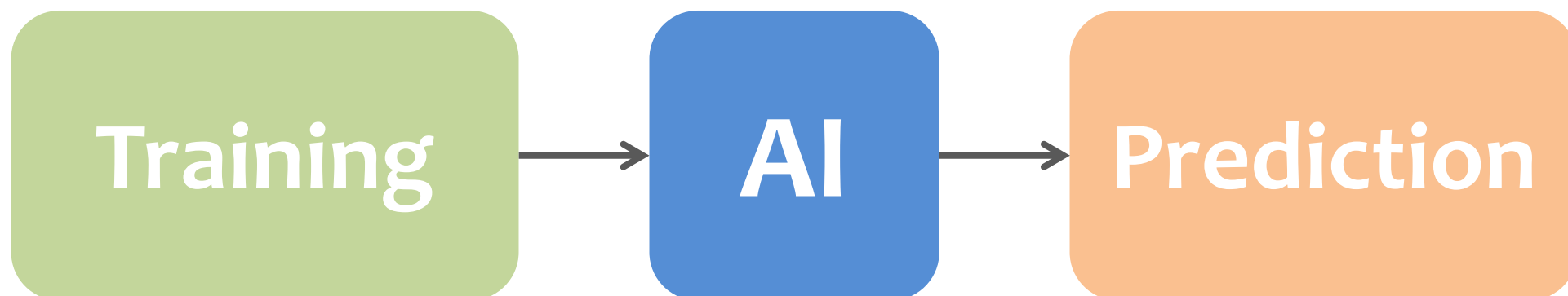
*Alexandros Labrinidis, H. V. Jagadish:

Challenges and Opportunities with Big Data. [PVLDB 5\(12\)](#): 2032-2033 (2012)

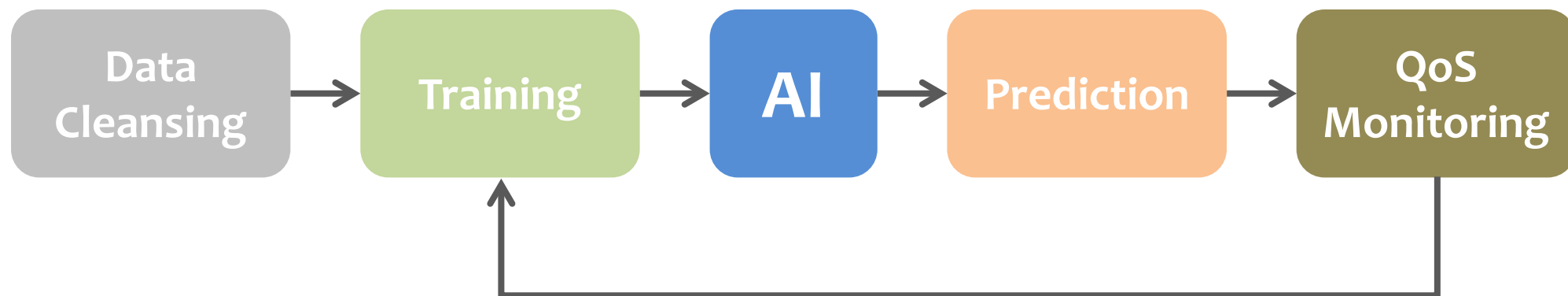
AI

AI

In the nutshell

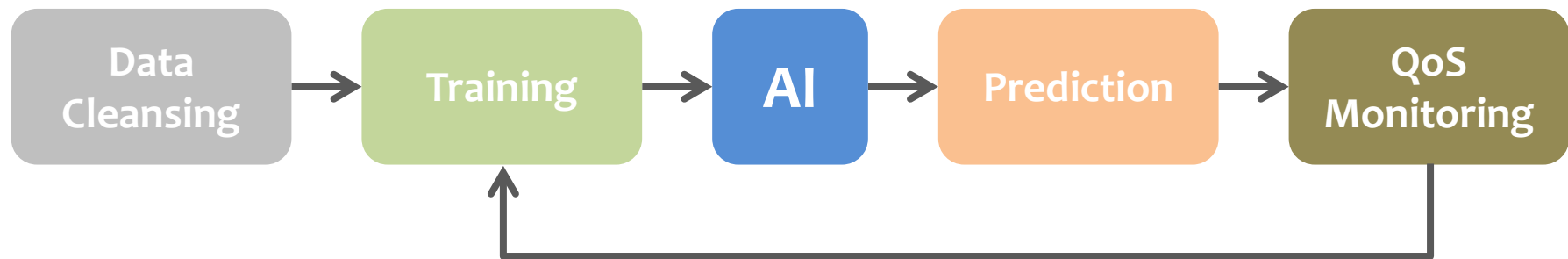


Preparation and QoS



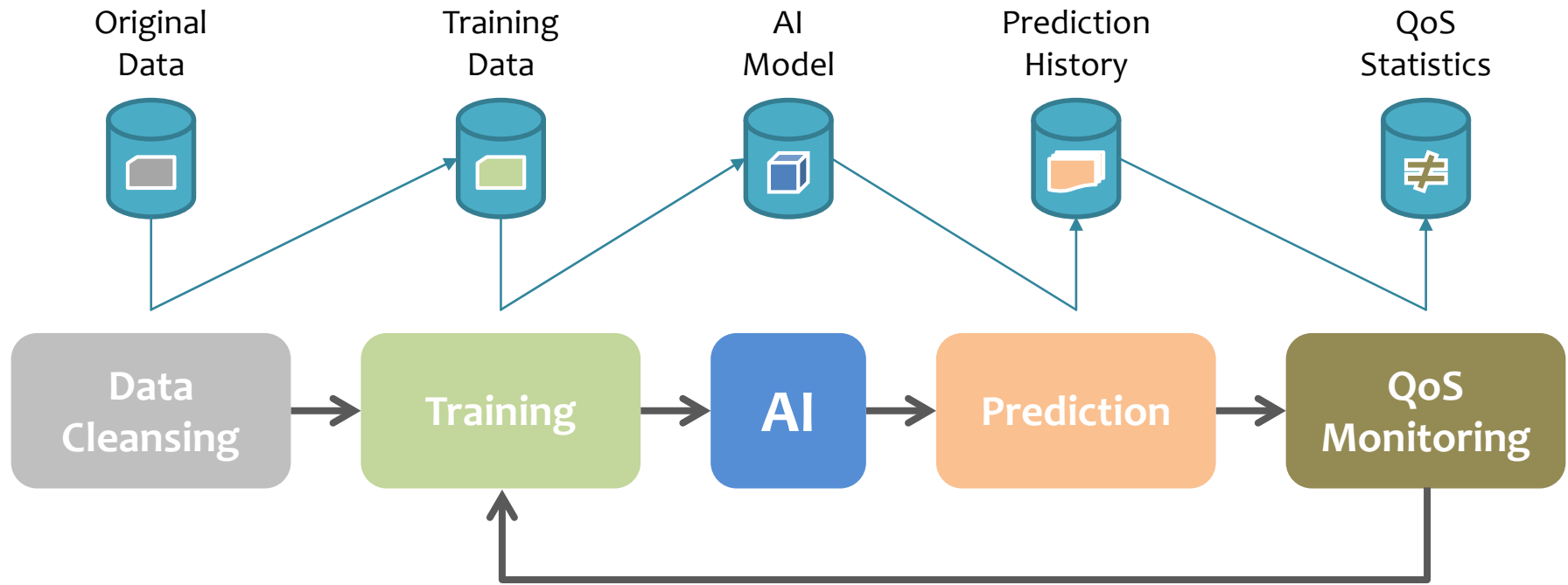
The Workflow

Data Collections



Query Interactions

Data Collections

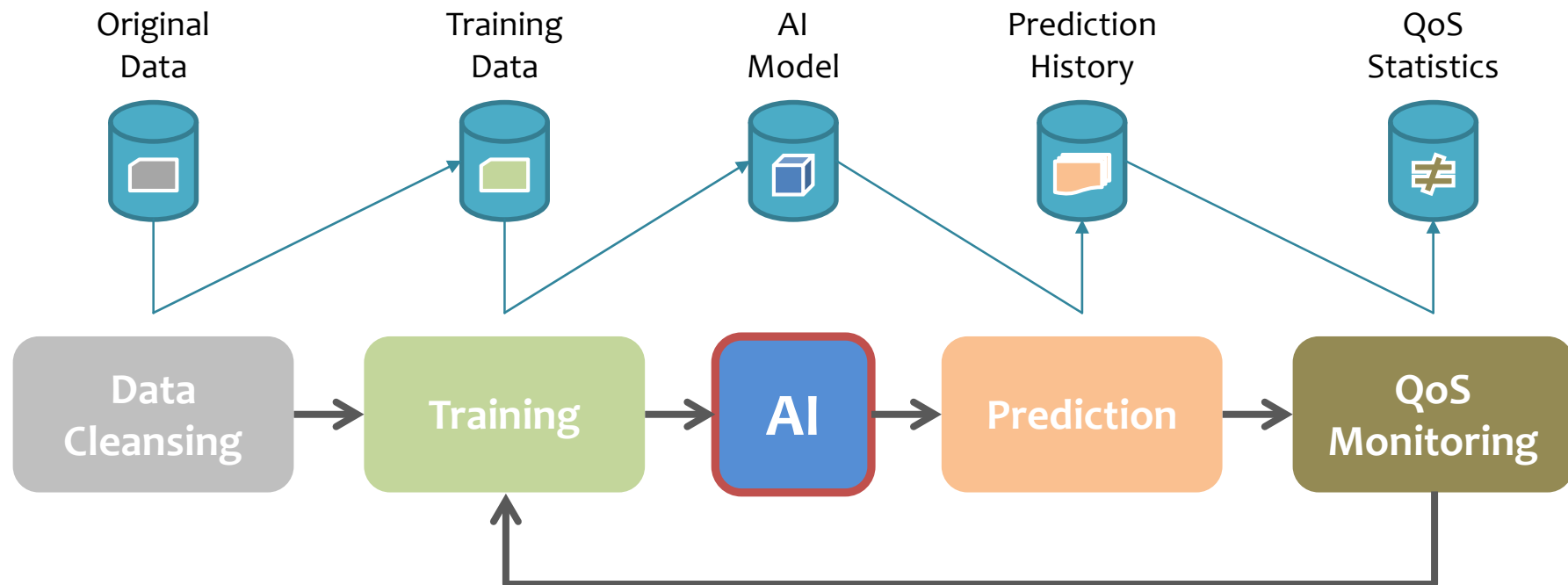


Query Interactions

Scaling AI

How to manage data for multi-user/application?

Data Collections



Query Interactions

How to coordinate
*collaborative data
cleansing?*

How to support
*machine/deep learning
under privacy concern?*

How to preform timely
*quality control for online
prediction?*

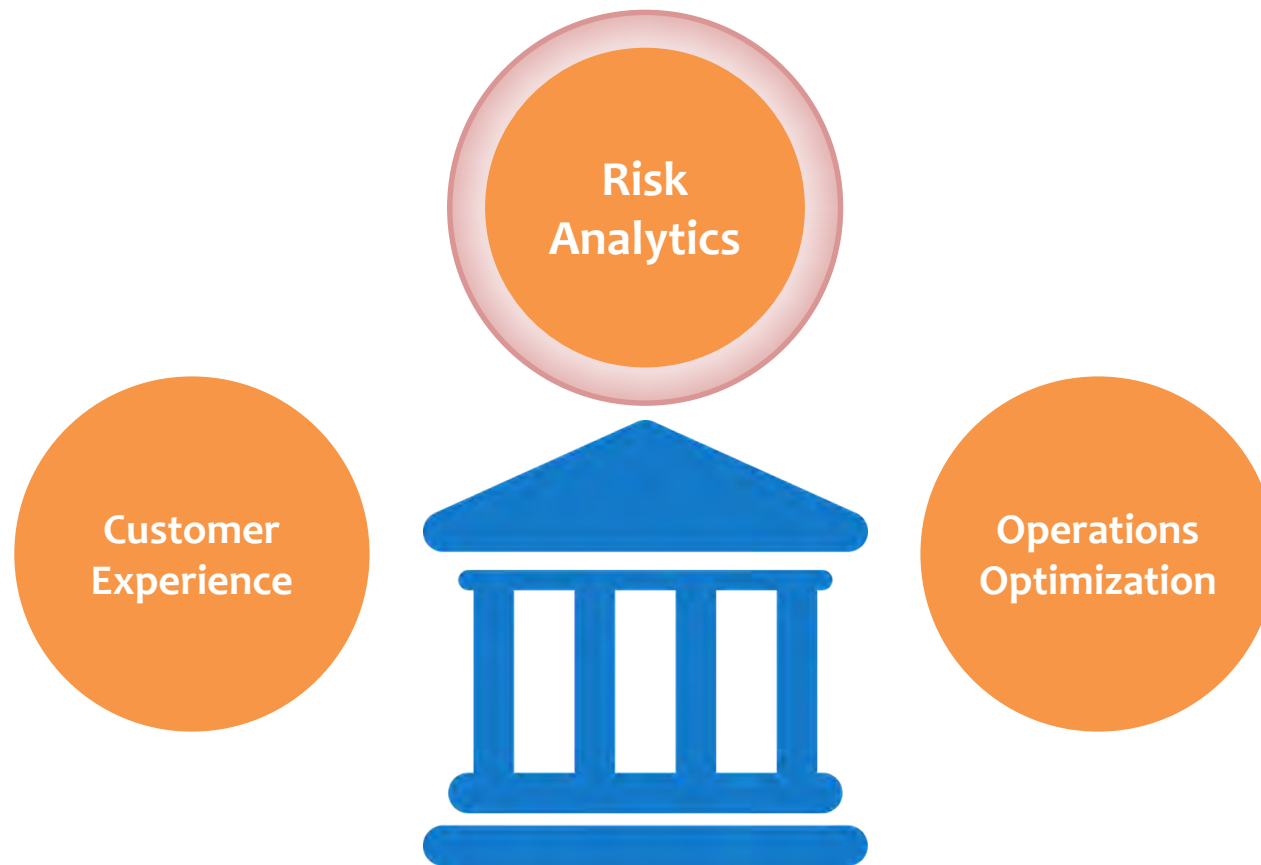
ForkCloud

GIT FOR DATA

Data and Learning

- Due to PDPA/GDPR, data is not widely available
- Even if it is stored in the cloud, data cannot be seen by external AI experts and developers
- Without data, models cannot be well designed or trained
!!
- No Free Lunch Theorem [1997]
 - *there is no useful model/algorithm for all the data distributions*

Big Data Analytics in Banking



Scattered
Data

No Single View of
the Customer

Fraud
Identification

Governance

Risk Analytics



- Customer basic information
- Customer asset status
- Customer transaction history
- Customer trading pattern (e.g., trading accounts, trading frequency, ...)
- Customer credit history
-

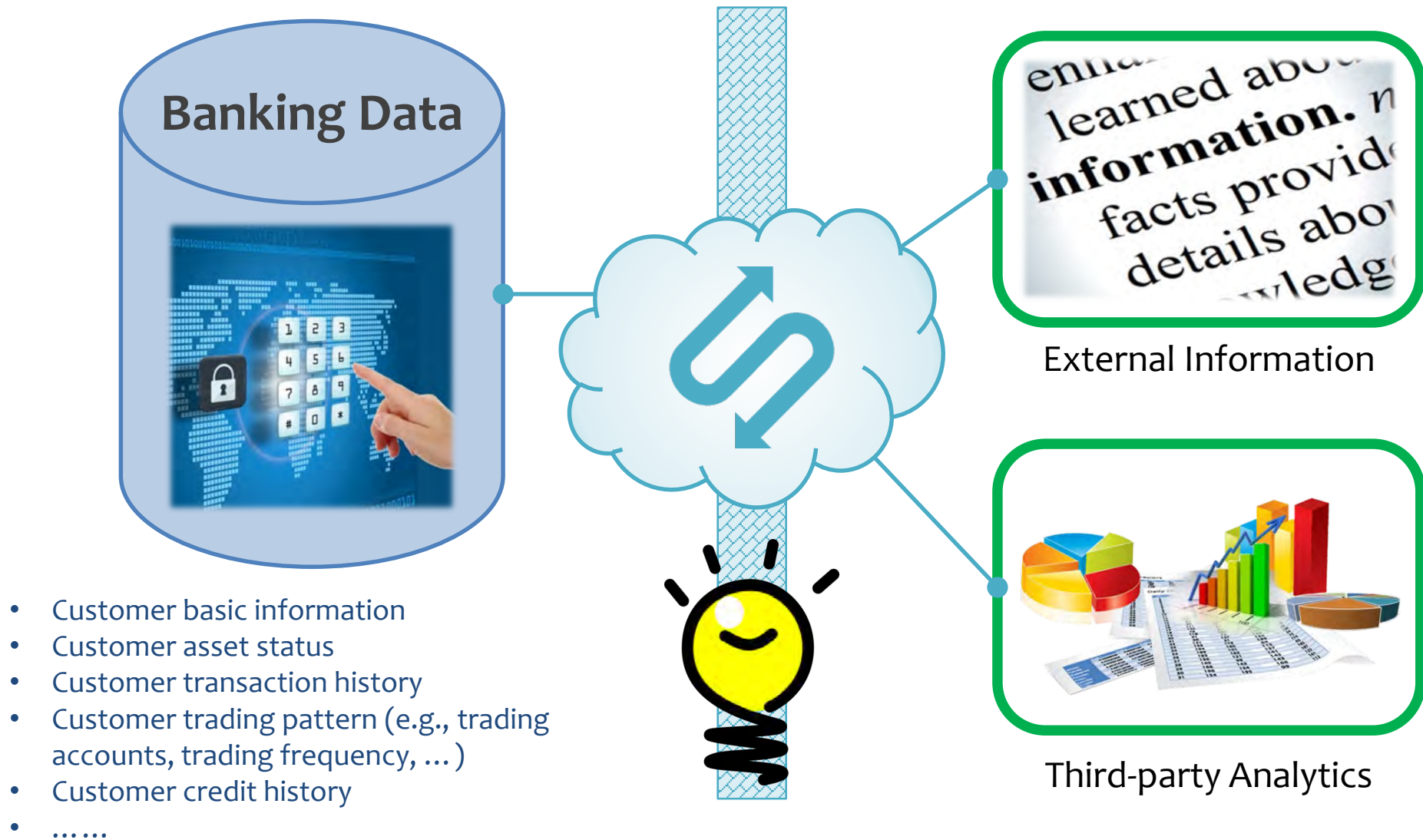


External Information



Third-party Analytics

Risk Analytics



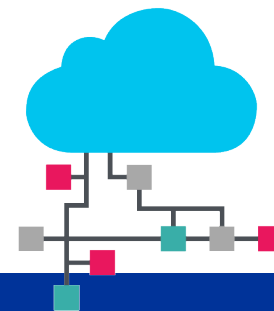
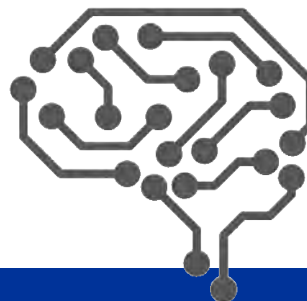
Facilitating AI Services

Git for data

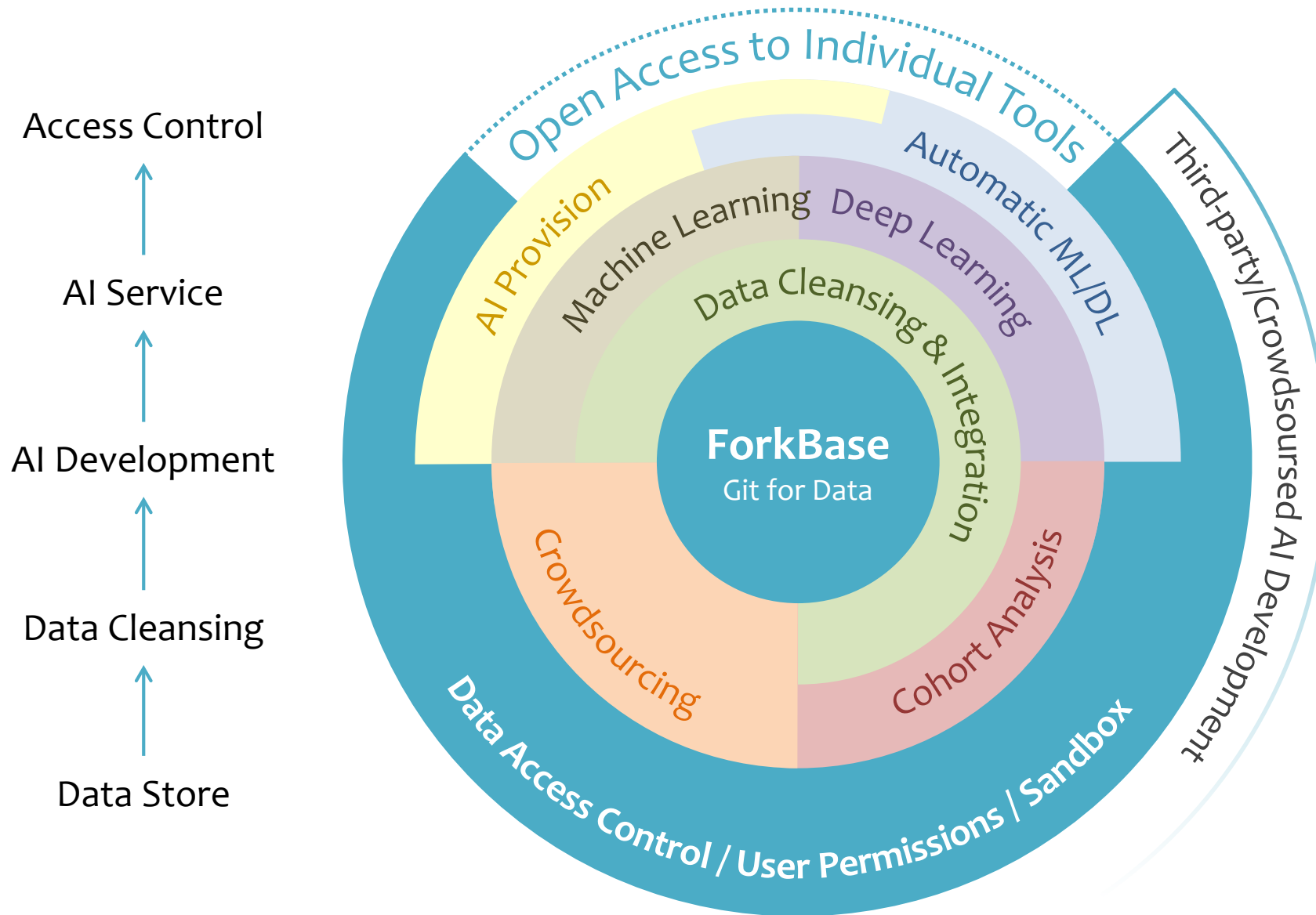
- Data-driven services to facilitate development of domain-specific AI solutions
- Data cleansing, curation and integration services
- Data versioning to facilitate collaboration, data provenance and improvement of quality data

Applications / Target market

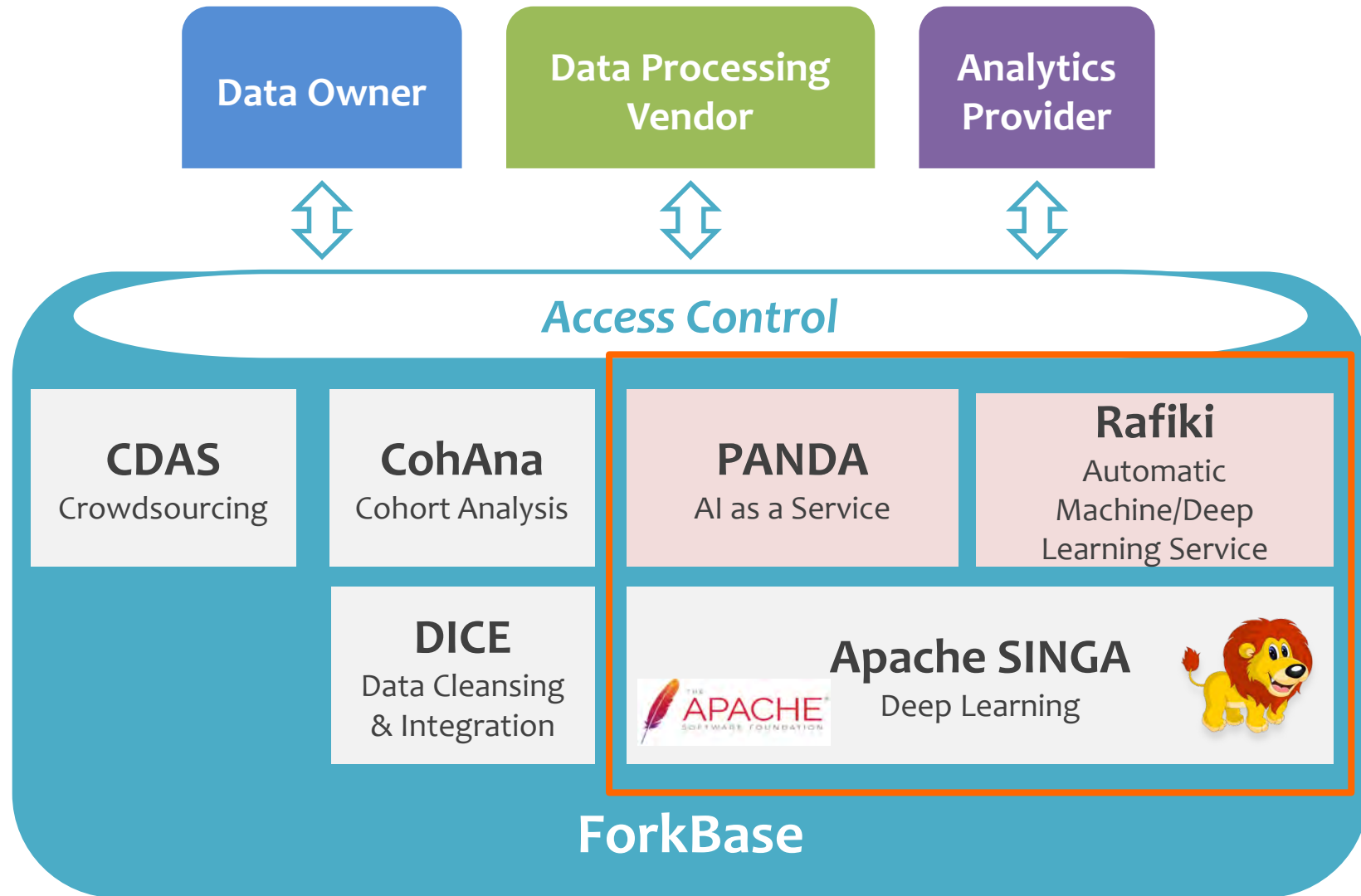
- “Git” for data – cleaning and curation, privacy-preserving crowdsourcing for service, data protection and provenance, data **immutability**
- AI-as-a-Service in the cloud
- Tools for data scientists and business analysts to conduct analytics on sensitive data

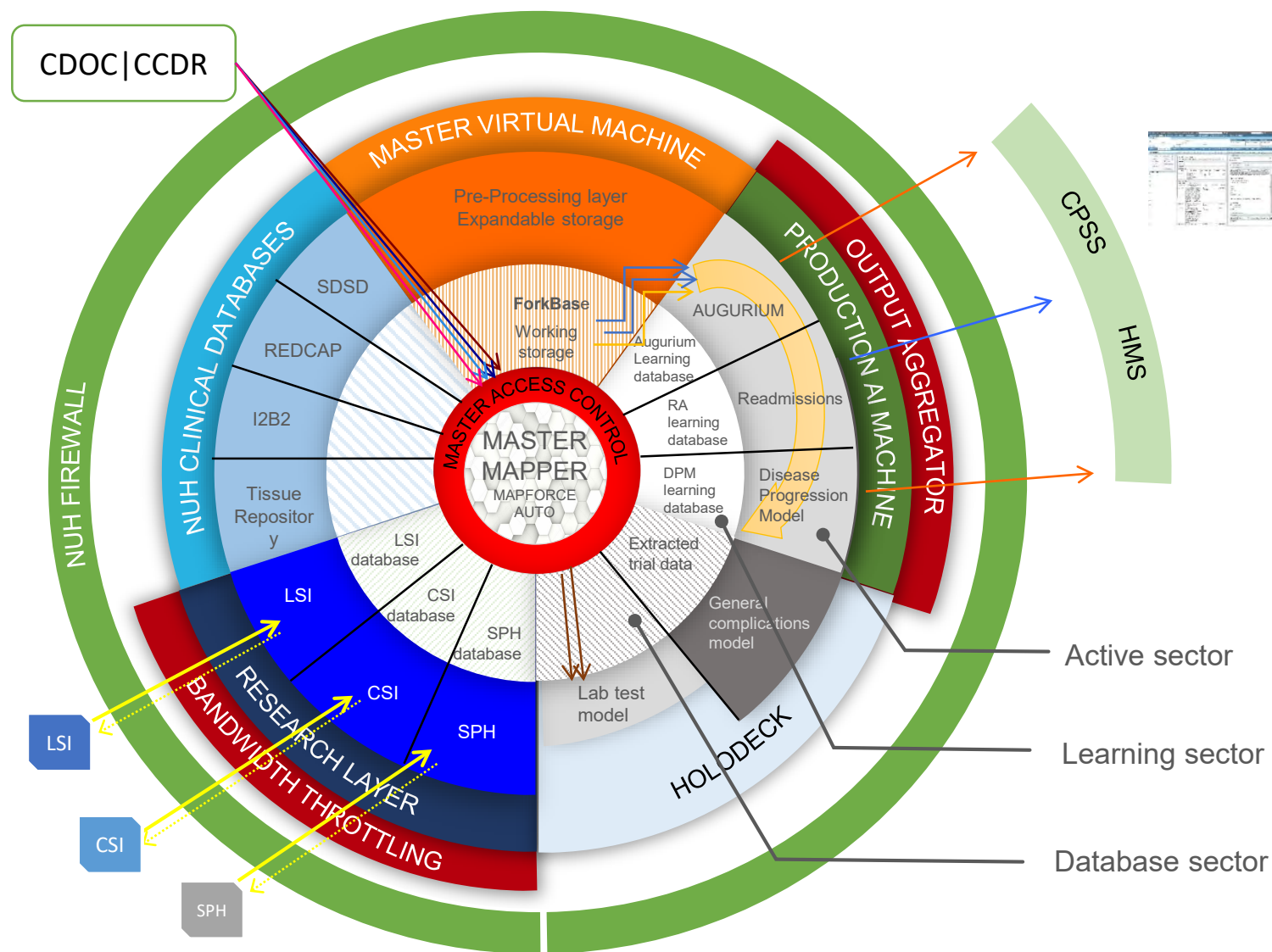


The ForkCloud Stack



Software Stack and Architecture





Financial News Analytics

For fintech applications

STOCK PRICE VS BREAKING NEWS

World Stock Market Performance

MSCI All Country World Index with Selected Headlines from 2012



Source: MSCI.

In US dollars. Index is not available for direct investment. Performance does not reflect the expenses associated with management of an actual portfolio. Past performance is not a guarantee of future results.

Risk

UK foreign office says it's exploring a 'sanctions regime' with the EU against Saudi Arabia

CNBC • today

- **Trump refuses to sanction Saudi crown prince over death of Khashoggi**

Daily Mail • today

- **Trump ally says Saudi sanctions must be in budget, as lawmakers rip Trump's defense of crown prince**

Washington Post • 3 hours ago

- **US Senate Demands Trump Explore Saudi Sanctions Over Khashoggi Murder**

Sputnik International • 3 hours ago



[This Photo](#) by Unknown author is licensed under [CC BY-SA](#).

Risk

IsoRay (ISR) Downgraded by Maxim Group to Hold

Fairfield Current • 3 days ago

Share

Cracker Barrel Old Country Store (CBRL) Lowered to Hold at Maxim Group

Baseball Daily News • today



Maxim Group Downgrades IsoRay, Inc. (ISR) to Hold

StreetInsider.com • 11 days ago

Opportunity

Huawei Mate Pro 20: What would a real mate do?

Irish Examiner • yesterday   

- **Huawei set to launch Mate 20 Pro in India on November 27**

Times Now • today



View more 

Huawei Mate 20 Pro with a 5G-ready chipset to be launched in India on 27 November

Firstpost • one hour ago



[This Photo](#) by Unknown author is licensed under [CC BY-SA-NC](#).

Opportunity

Unilever to acquire 75% of Italian personal care business Equilibra

WebWire • 4 months ago



Unilever To Acquire 75% Stake In Italian Personal Care Firm Equilibra | ESM Magazine

ESM - The European Supermarket Magazine • 4 months ago



How Unilever Could Benefit From Acquiring Equilibra

Forbes Now • 4 months ago



TOO MANY NEWS ARTICLES/PUBLISHERS

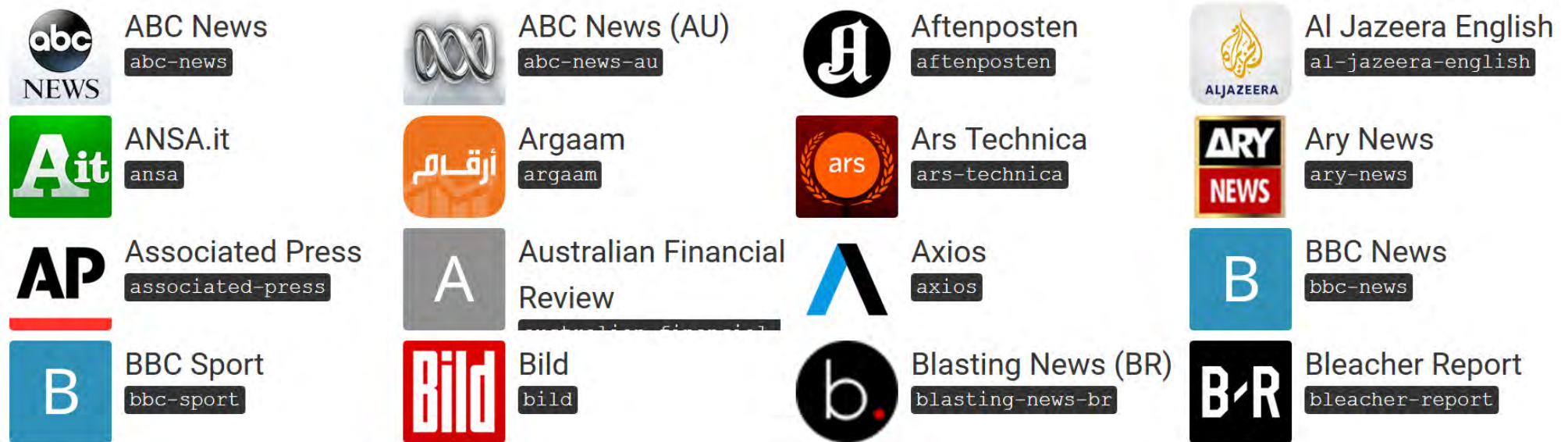


Image from <https://newsapi.org/sources>

Over 30,000 news publishers indexed by newsapi.org

Many articles



- **China says certain countries trying to force views on others at recently-concluded Apec summit**
The Straits Times • yesterday
- **China's former trade negotiator questions tariff strategy, Latest World News**
The New Paper • yesterday
- **US renews unfair trade accusations against China**
Channel NewsAsia • 4 hours ago
- **China turns focus on US with claim its Apec amendment was vetoed**
South China Morning Post • today



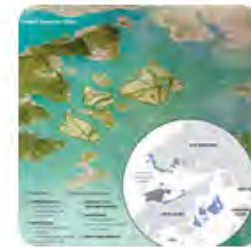
Few articles



Hong Lam Marine

Why the Lantau Tomorrow Vision needs a rebranding

South China Morning Post • 15 days ago • Opinion



Stop Lantau reclamation, say dolphin activists, after mega-bridge marine park deemed 'useless'

Hong Kong Free Press • 28 days ago



Sentek Marine Named Singapore's Biggest Bunker Supplier by Volume in 2017

Ship & Bunker • 10 months ago

Tasks



News deduplication



Knowledge graph construction



Risk/opportunity discovery



Stock price/index prediction

CLUSTER

1

RESEARCH REPORTS Maxim Group Downgrades Apple on Its Shift to Services

Today, 19:40 by admin Risk

These reports, excerpted and edited by Barron's, were issued recently by investment and research firms. The reports are a sampling of analysts' thinking; they should not be considered the views or recommendations of Barron's. Some of the

Apple gets downgraded on concern services focus won't deliver big profits

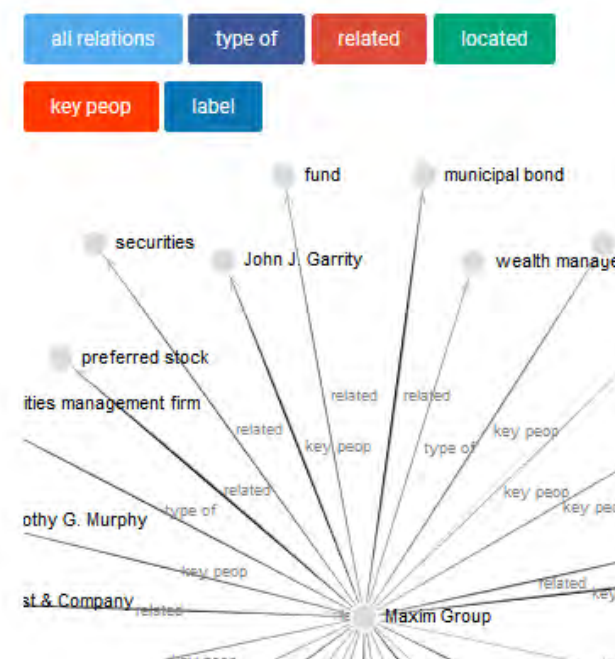
Today, 10:59 by admin

2

Risk/opportunity prediction

Knowledge graph if few articles

Visualization



News deduplication by clustering

- **Document Clustering**
 - *Extract representation of the article/document*
 - *TF.IDF, word2vector -> Document vector, etc.*
 - *Apply clustering algorithms*
 - *K-means, DBSCAN, etc.*
- **Challenges**
 - *Too many articles are generated every day -> Big data*
 - *Clustering algorithms are expensive*
- **Hierarchical clustering**
 - *Identify articles of the same company*
 - *Apply clustering algorithms over the articles about the same company*



Knowledge Graph

- **Knowledge**

- **Relation tuples, (A, relation, B).**

- *Tim Cook, CEO of Apple, has attended the meeting. → (Tim Cook, CEO, Apple)*
- *Foxconn that makes Apple's iPhones... → (Foxconn, Supplier, Apple)*

- **Application**

- **News extrapolation**

- *(Person A, CEO, Company B), (Company B, Supplier, Company C)*
- *If the big company C has risks, then the small company B is likely to have troubles*

- **Stock price/index analysis**

- *The drop of price for Company C would affect that of Company B*

Knowledge Graph Construction

- **Subtasks**
 - **Named Entity Recognition**
 - Locate and classify **named entity** mentioned in unstructured text into pre-defined categories, such as person, organizations, locations...
 - e.g., [John]_{person} bought 500 shares of [Apple Inc.]_{organization} In [2015]_{time}.
 - **Entity Linking**
 - Link the mention into the entity in the knowledge base.
 - e.g., "Paris is the capital of France" → Paris is referring to this entity.
 - **Relation Prediction**
 - Classify the semantic relationship between entities
 - e.g., "Yesterday, Foo Inc. announced their acquisition of Bar Corp." -> (Foo Inc., acquire, Bar Corp)

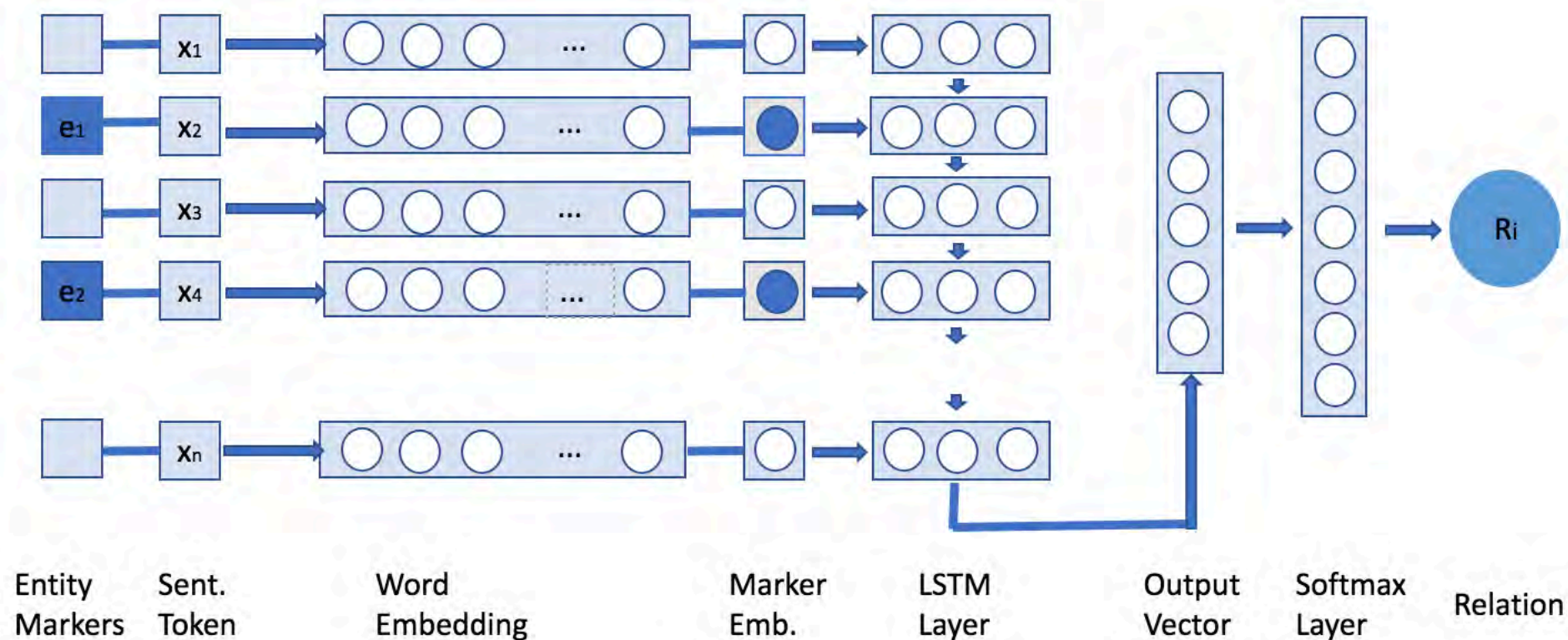
Knowledge Graph Construction

- **NLP is challenging**
 - *Ambiguous words*
 - *e.g., Apple & apple*
 - *Uppercase words impacts*
 - *e.g. A and B Jointly Developed ... "B Jointly Developed" can be mistakenly recognized as one entity*
 - *Multiple relations in one sentence*
 - *e.g., Mary married with a guy, who is the chairman of Bar Corp.*

Knowledge Extraction --- Deep Learning

- **Supervised learning**
 - *Select the interested relations*
 - *Key Person (e.g., employee, founder, CEO, chairman, head, etc)*
 - *Related Company (e.g., supplier, partner, competitor, etc)*
 - *Location*
 - *Select relevant entities*
 - *PERSON, LOCATION, ORGANIZATION*
 - *Train neural networks that map (sentence, entities) -> relation*
 - *Bill Gates is the founder of Microsoft. -> (Bill Gates, founder, Microsoft)*
 - *Mondrian takes 6.6% stake in Sheng Siong (Mondrian, shareholder, Sheng Siong)*

Knowledge Extraction --- DL Data Flow



Knowledge Extraction --- Distant Supervision

- **Steps**
 - **Find text (sentences) discussing the entities of a relation**
 - *Zuckerberg says he won't step down as Facebook CEO*
 - *Apple's Tim Cook says tech regulation 'inevitable'...*
 - *not talking about the CEO relation → noisy sample*
 - **Learn over the noisy text to correlate**
 - *the descriptions, e.g., step down, CEO, Facebook*
 - *the relation, CEO*
 - **Infer the relation of entities in new sentences**
 - *Microsoft CEO Satya Nadella sells 30 percent of common stock*
 - *Returns: Satya Nadella, CEO, Microsoft*

Risk/opportunity detection

- **Sentiment analysis**

- Too general

- **Risk**

- Rating downgrade
- Sanctions
- More

- **Opportunity**

- New market
- New product
- More

text	label
Unilever to increase investment in Greece	New market/opportunity
Japan cryptocurrency exchange plans Europe, Asia sales hubs	New market/opportunity
Trump administration unveils sanctions aimed at starving North Korea of resources	Sanctions/risk
Why Did Maxim Group Downgrade Domino' Pizza?	Rating downgrade/risk

Risk/opportunity detection

- **Document classification**
 - *Annotate the articles as training data*
 - *Expensive*
 - *Inconsistency between annotators*
 - *Data cleaning and pre-processing is important*
 - *Cluster articles for fast annotate*
 - *Rule-based annotation*
 - *...*
 - *Deep learning models for classification*
 - *CNN, RNN*
 - *BERT (bidirectional encoder representations from transformers)*

Stock Prediction

- Traditional solutions for stock prediction are based on time-series models.
- With the recent success of deep neural networks in modeling sequential data, deep learning has become a promising choice for stock prediction.

Related Work

	Topic	Type of Data	Number of stocks
Temporal Relational Ranking for Stock Prediction - Feng et al (2019)	Stock selection	Stock Prices	1700
A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction – Yao et al (2017)	Index Prediction	Stock Prices	100
Deep Learning for Event-Driven Stock Prediction - Xiao et al (2015)	Index Prediction	News and Stock Prices	500
Stock price prediction via discovering multi-frequency trading patterns – Zhang et al (2017)	Index Prediction	Stock Prices	50

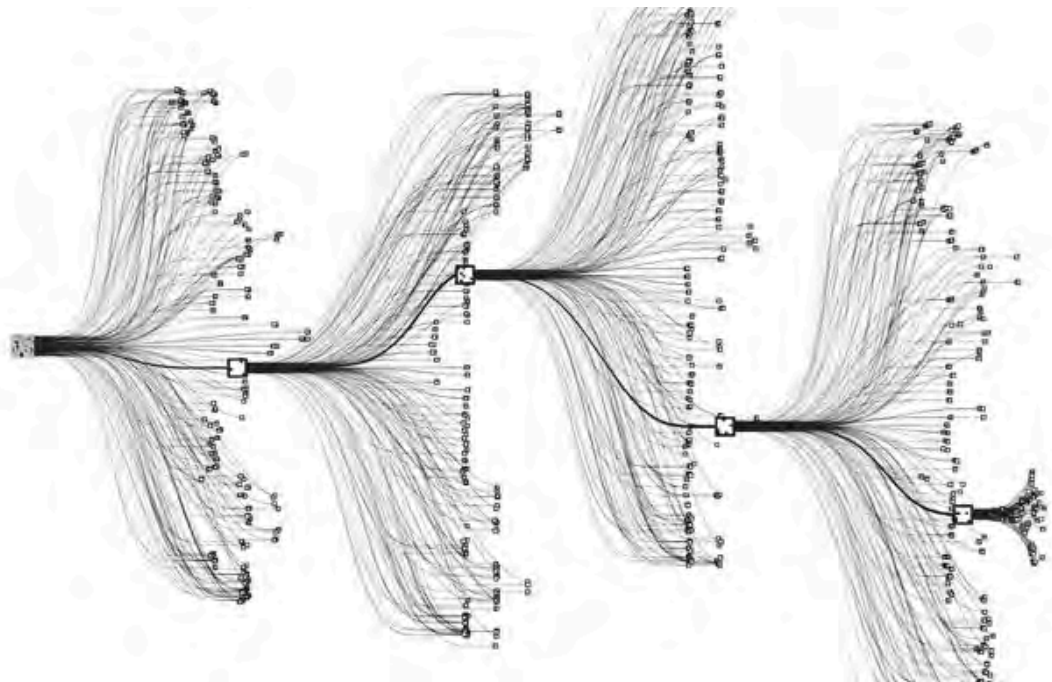
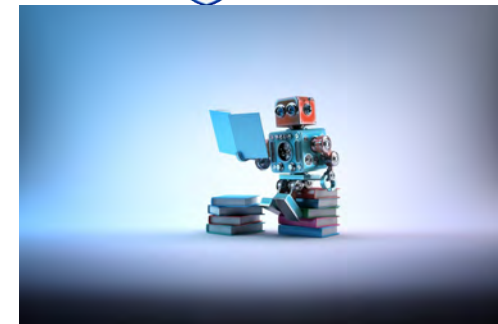
Optimized Trade Execution

The goal: is to sell (respectively, buy) V shares of a given stock within a fixed time period (or horizon) H , in a manner that maximizes the revenue received (respectively, minimizes the capital spent).

Trade-offs: such as between the speed of execution and the prices obtained for the shares.

Virtually every strategy's profitability will depend on how well it is implemented.

Methodology



Try all possible
actions!

Experiment results

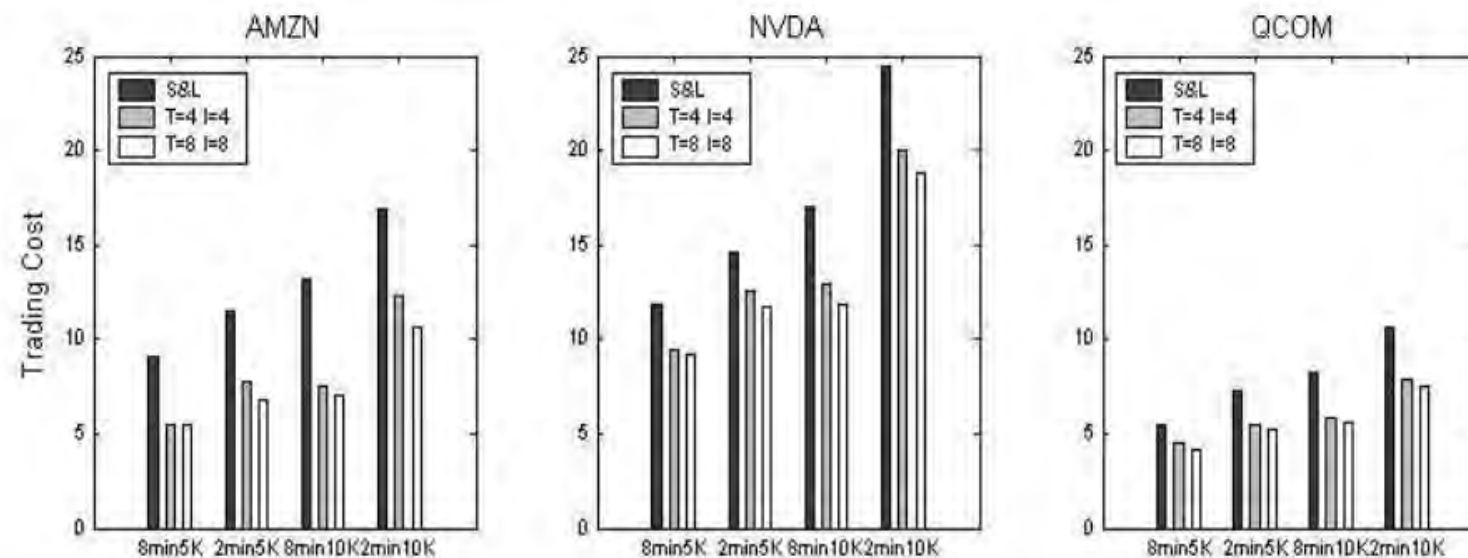


Figure 3. Expected cost under S&L and RL: adding private variables T and I decreases costs

Fraud Detection

Deep Learning Opportunity



Current models can only catch ~70% of all fraud cases



Traditional ML models view transactions atomically



Often missed fraud transactions are part of a series



Capturing correlation across many features

Summary

- AI has provided great improvement in prediction on a variety of applications, including high stack applications
- However, unlike games, machines cannot learn by playing out every possibility, and also, there are just too many external factors
- AutoML, NAS, **Model Verification** have reduced the barrier for adoption, but a lot more needs to be done
- Data is a new oil, and data quality determines its value but, sensitive data is problematic to handle
- **GIT for Data** for data provenance and governance, and supporting AI development without disclosing the data

Thanks!