



AI and Machine Learning in Finance

Maxime Tchibozo

International FinTech Program

HEC Paris

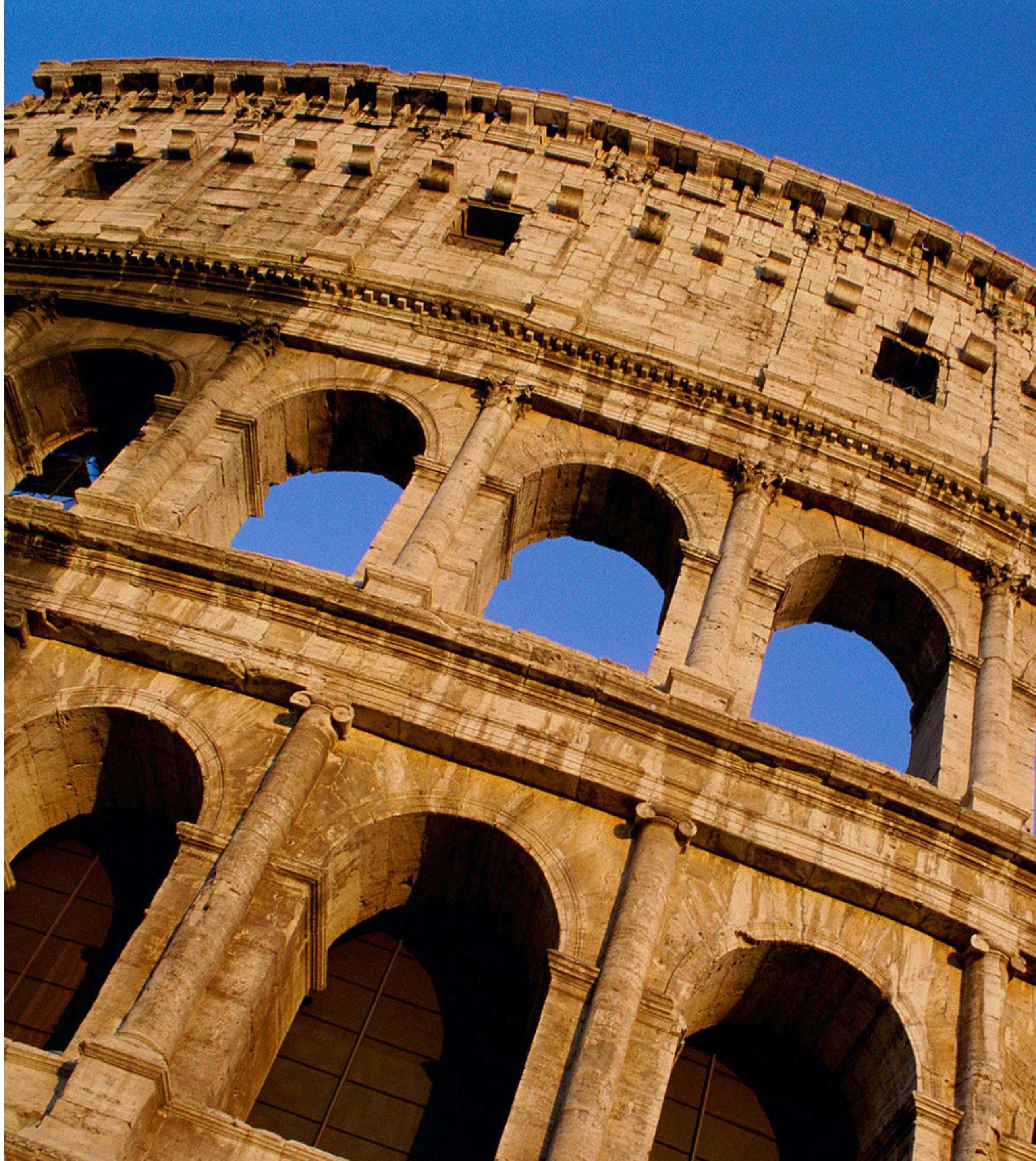
June 2022

Agenda

- Why are Financial Institutions interested in AI / ML?
 - Overview of core AI concepts
 - Case Study 1: Trading
 - Case Study 2: Investment Banking
 - Frontiers of AI in Finance
 - Risks for AI Adoption in Finance
 - Final Thoughts
- This is a **fully collaborative** class.
Please raise your hand, **introduce yourselves & tell us** about your prior experiences.

Why are Financial Institutions interested in AI/ML?

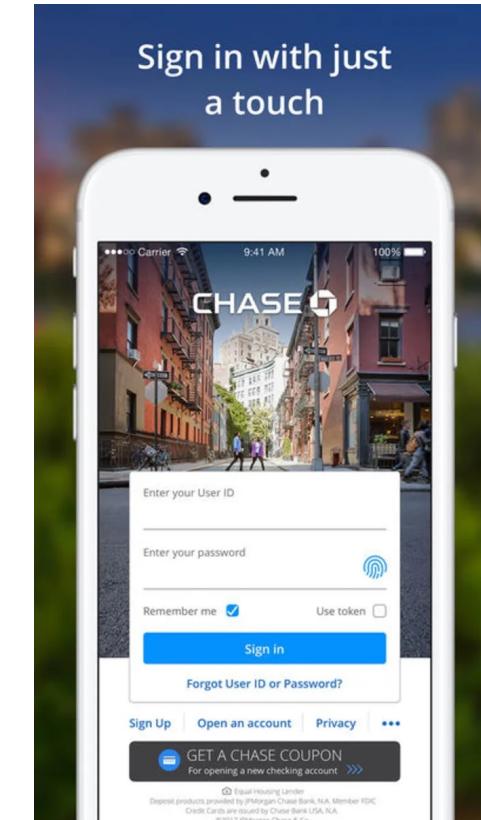
- Financial Institutions are Investing Heavily in AI
- Why Now? A Timeline of AI Adoption
- Topics in AI



Financial Institutions are Investing Heavily in AI

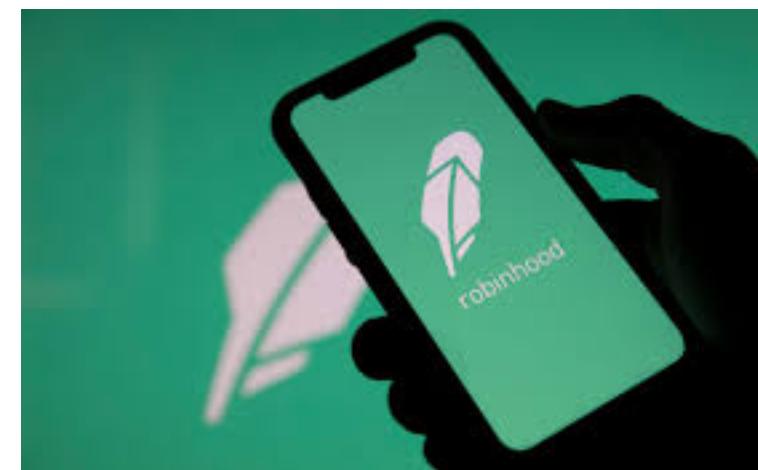


aladdin®
by BLACKROCK®



Shift Technology

LA DÉTECTION DE LA FRAUDE À L'ASSURANCE LES MATHÉMATIQUES DU BIG DATA L'INFORMATIQUE EN CLOUD ET LE SAAS



- Investment Banks & Asset Managers
- Retail Banks
- Insurance Companies
- Brokerage Firms
- Government Entities

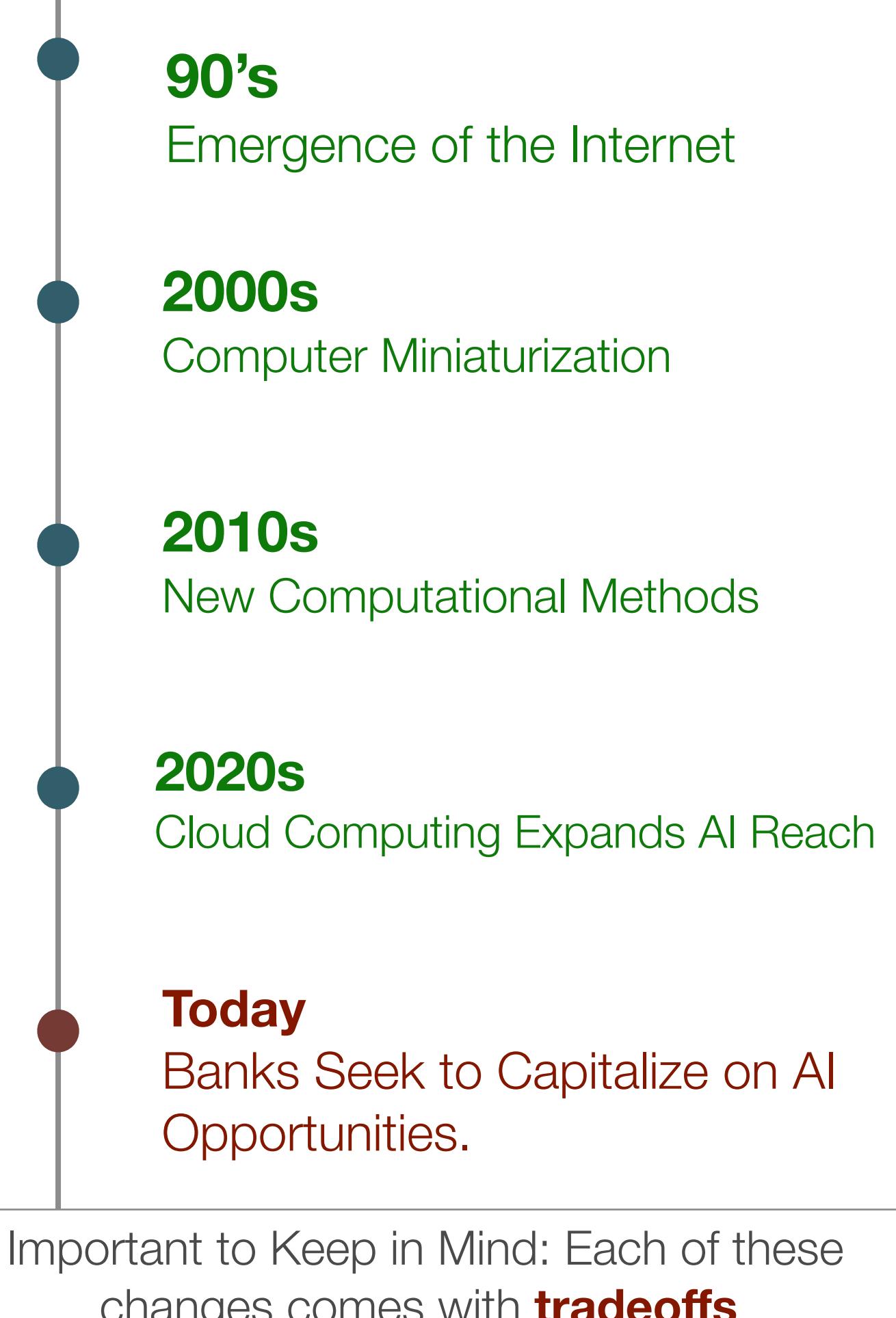
Why Now? How the Current Paradigm of Data Appeared

The **AI Paradigm** we see today was born from big investments Tech companies made into the Digital world.

Tech companies made these investments to pursue the **fastest possible growth**, valuing scaling above all else.

Banks have **vastly different priorities**. Understanding these differences allows us to find exploitable opportunities.

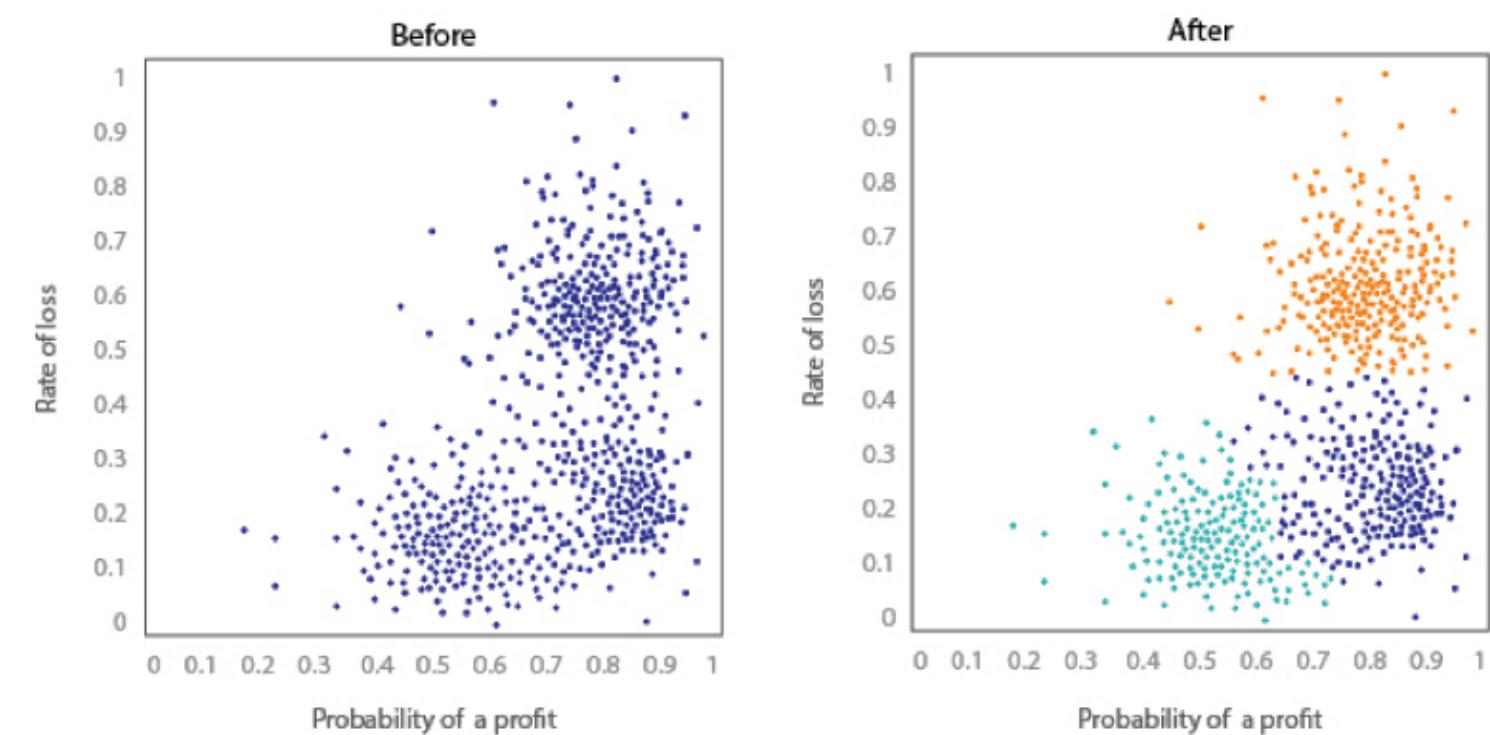
How Banks will use AI is question that will get solved over the coming years.
Today, all ideas are welcome.



Topics in AI

Unsupervised Learning

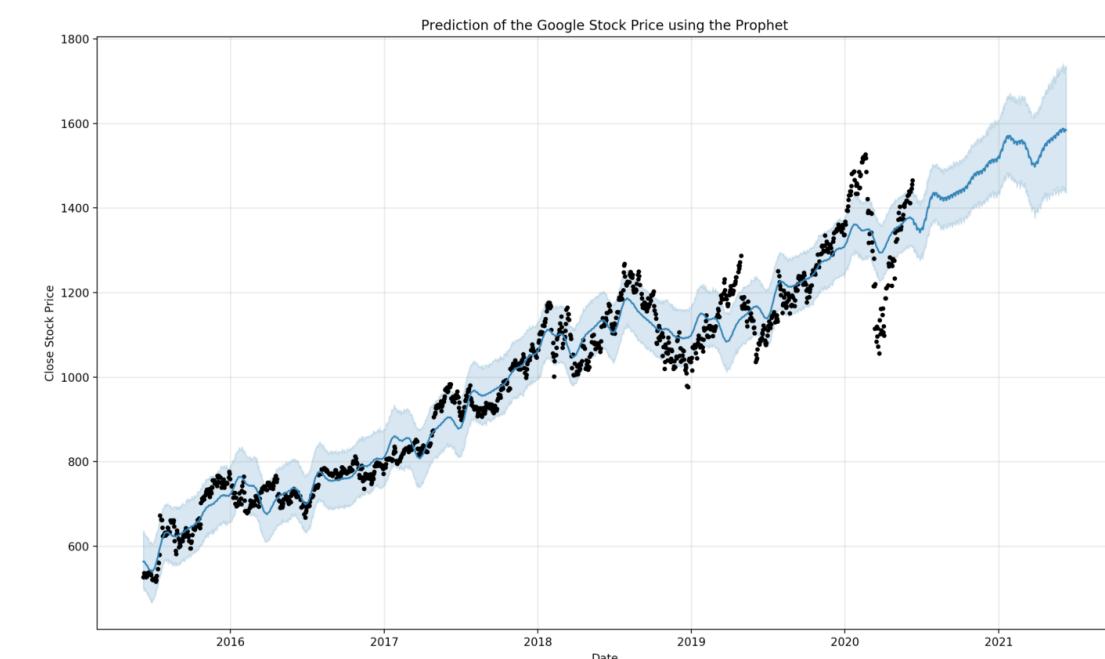
Find patterns in data without human guidance



“Here is all the data we have on past trades, can you find an interesting way to group similar trades together”

Supervised Learning

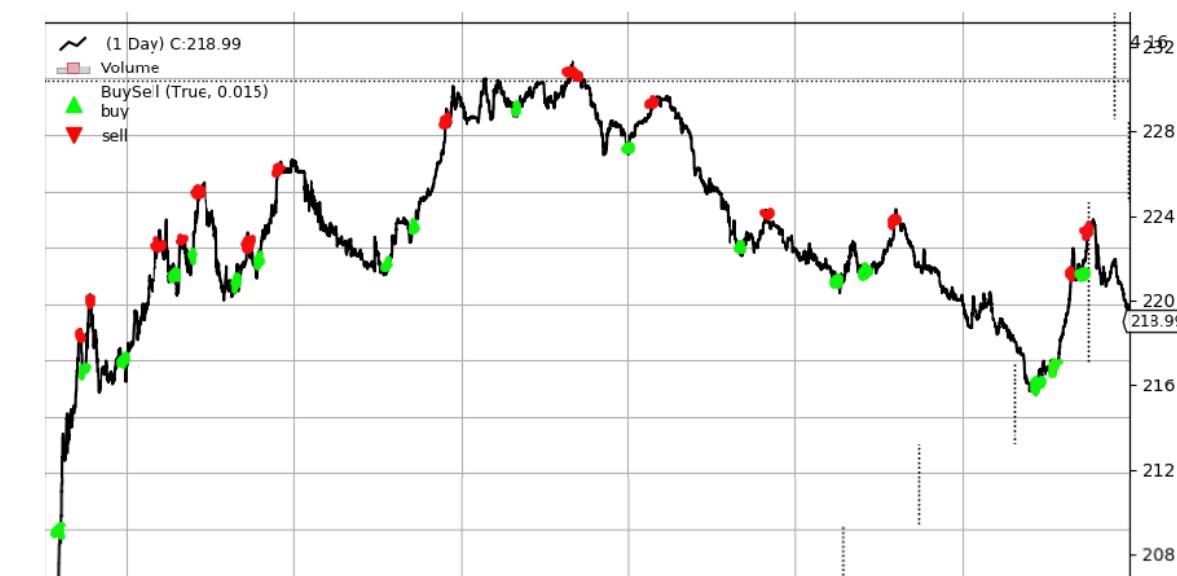
Learn patterns which best replicate a predefined outcome



“This is what the market looked like each time we made a trade up till now. *Based on those past trends*, can you tell me how the price could evolve after our trade today?”

Reinforcement Learning

Learn to **apply** optimal **policies** in a particular environment given specific reward incentives



“Can you tell me in real-time when you think it's the right moment to buy or sell?”

Topics in AI

Unsupervised Learning

Find patterns in data without human guidance

- Clustering & Segmentation
- Recommendation Systems
- Compression & Representation Learning
- Analytics & Visualization
- Data Generation & Augmentation

Supervised Learning

Learn patterns which best replicate a predefined outcome

Reinforcement Learning

Learn to **apply** optimal **policies** in a particular environment given specific reward incentives

Topics in AI

Unsupervised Learning

Find patterns in data without human guidance

Supervised Learning

Learn patterns which best replicate a predefined outcome

Reinforcement Learning

Learn to **apply** optimal **policies** in a particular environment given specific reward incentives

- Computer Vision & Natural Language Classification
- Object Detection & Entity Recognition
- Forecasting & Predictive Modeling

Topics in AI

Unsupervised Learning

Find patterns in data without human guidance

- Real-Time Decision Making
- Genetic Algorithms

Supervised Learning

Learn patterns which best replicate a predefined outcome

Reinforcement Learning

Learn to **apply** optimal **policies** in a particular environment given specific reward incentives

Overview of core AI Concepts

- An AI Example: Setting Mortgage Interest Rates
- Identifying Main Factors
- A Data Science Approach
- Data & Representations
- Metrics & Evaluation
- Model Criticism & Update



An AI Example: Setting Mortgage Interest Rates

How should a Bank set rates for customers looking to buy a home?

A **mortgage** is a type of loan used to purchase real estate. The borrower agrees to pay the lender - usually a Bank - a series of recurring payments covering the principal amount of the loan and the interest.

The interest is referred to as the **mortgage rate**.

Which AI **topic** could this example fit under?

Which **factors might affect** how a Bank sets mortgage rates?



In France in 2022, the typical mortgage rate range is **1.2% - 2.5%**.

Setting a Mortgage Interest Rate - Main Factors

Personal Factors

Credit score, down payment & collateral, type of property, buyer profile, ...

Market Factors

Central Bank interest rates, inflation, competing banks, bonds, ...

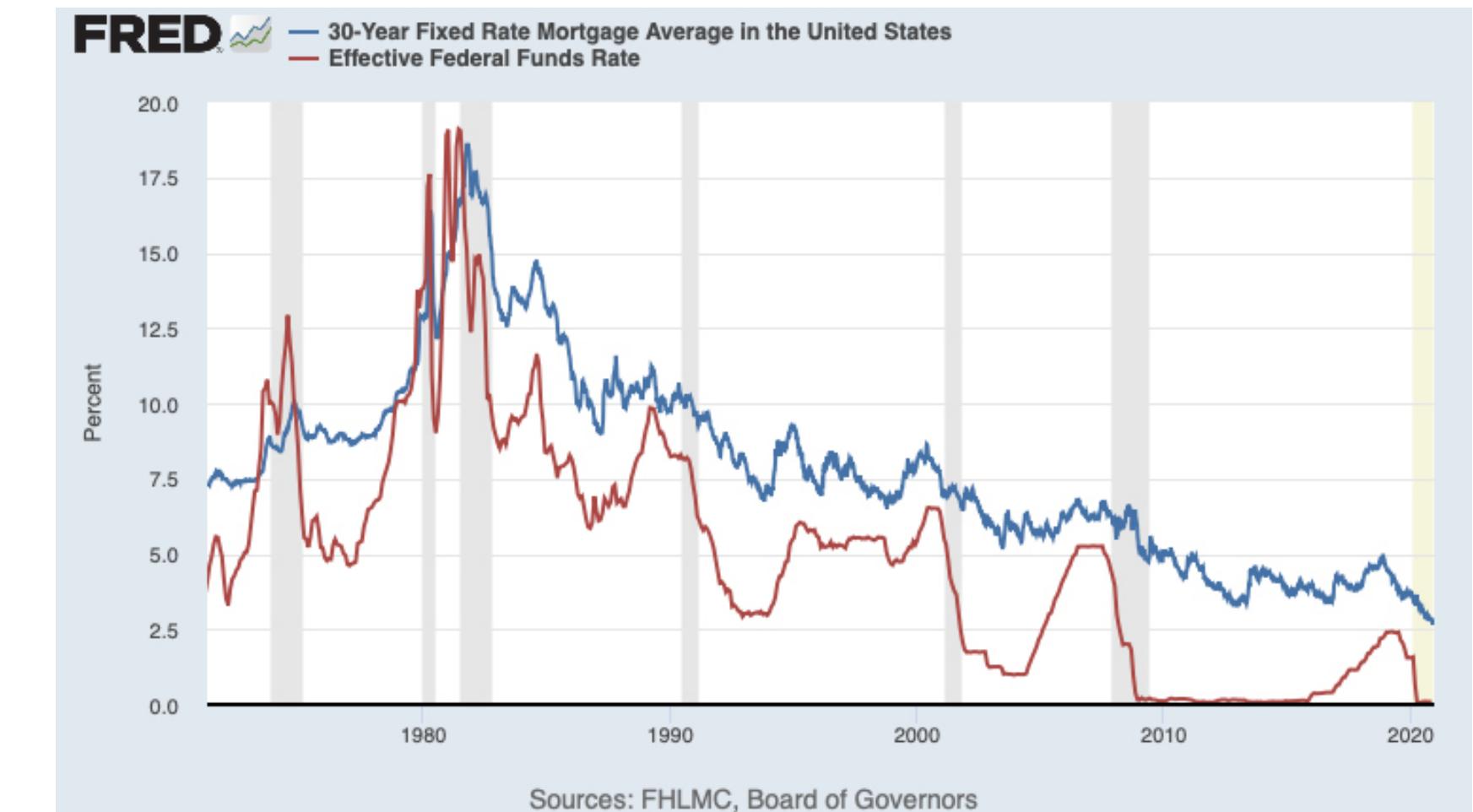
Product-related Factors

Loan maturity, loan amount, ...

Key Considerations

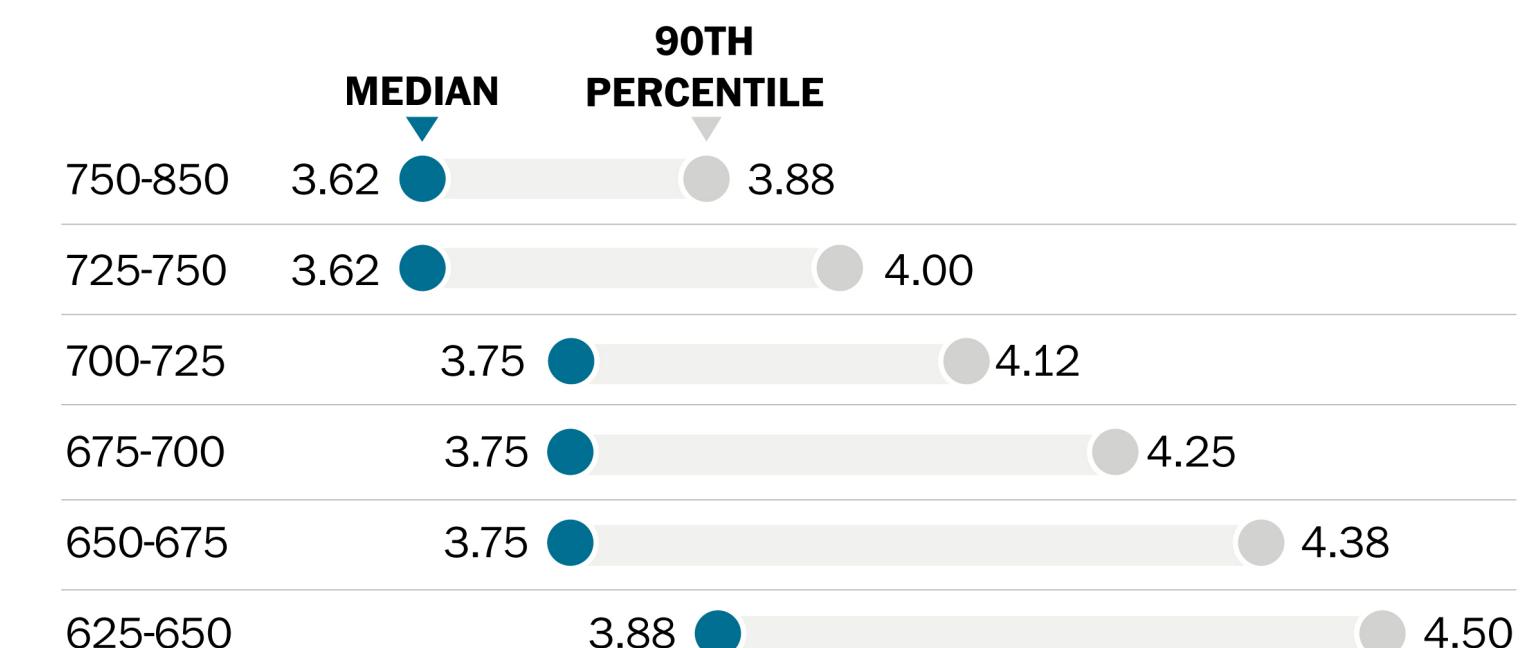
How does each factor **affect** the Bank's profit (all else held equal)?

How should each of these factors be **weighed**?



Higher credit score, lower mortgage rate

People with higher credit scores receive lower mortgage rates. While the median mortgage rate does not change with every bump in credit score, mortgage rates at the 90th percentile improve as scores rise.

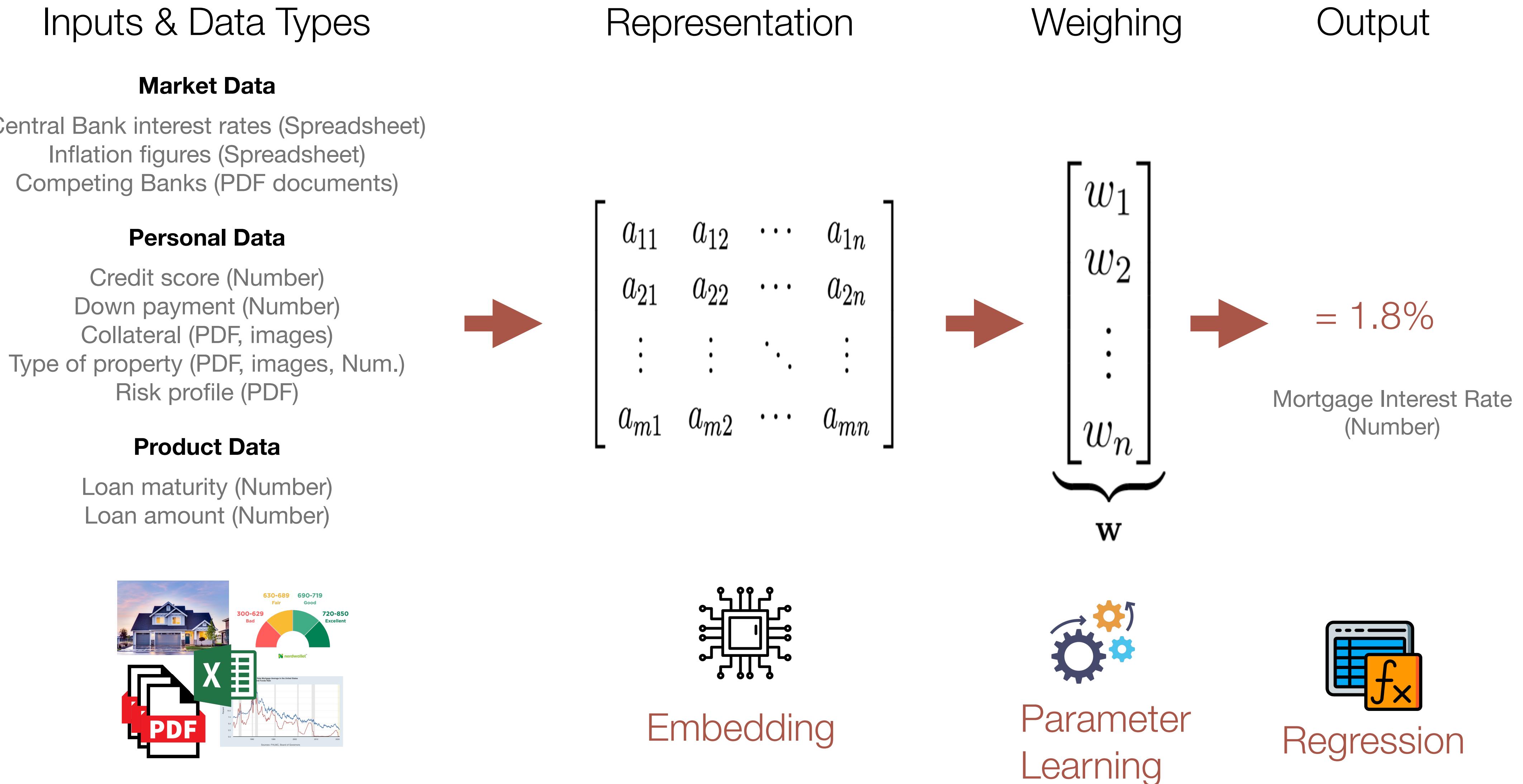


Note: Rates are for 30-year fixed-rate mortgages processed from Sept. 1 through Nov. 7 2016.

Source: Realtor.com

THE WASHINGTON POST

Setting a Mortgage Interest Rate - Data Science Approach



Setting a Mortgage Interest Rate - Data & Representations

Inputs & Data Types

Market Data

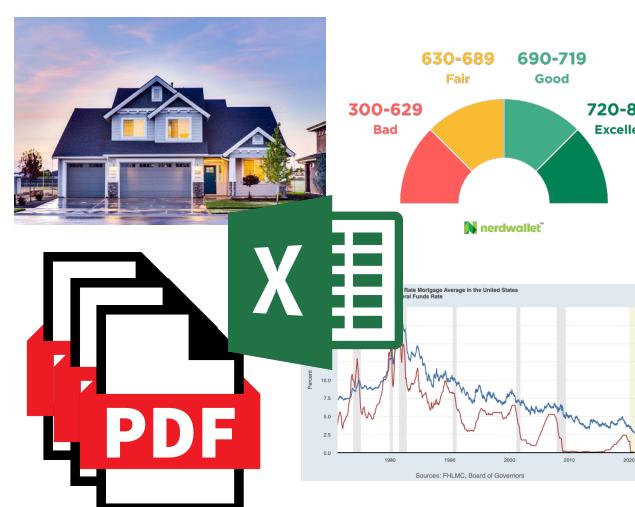
Central Bank interest rates (Spreadsheet)
Inflation figures (Spreadsheet)
Competing Banks (PDF documents)

Personal Data

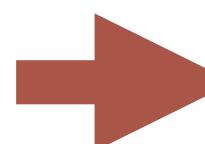
Credit score (Number)
Down payment (Number)
Collateral (PDF, images)
Type of property (PDF, images, Num.)
Risk profile (PDF)

Product Data

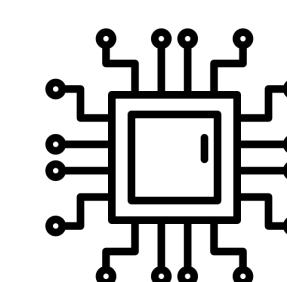
Loan maturity (Number)
Loan amount (Number)



Representation



$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$



Embedding

All data is preprocessed and converted into a **representation** (a.k.a **embedding**).

A good embedding **compresses information** while **preserving** the most important **properties** of the data.

AI algorithms learn patterns from embeddings directly.

Careful consideration is required:

- Raw inputs have vastly different formats
- Numeric values might be scaled differently
- Documents might contain mostly irrelevant information
- Data may not cover all the edge cases that appear in reality

Setting a Mortgage Interest Rate - Metrics & Evaluation

AI models transform embeddings into predictions.

Based on how close predictions are to reality, an algorithm will update the weights and the embeddings.

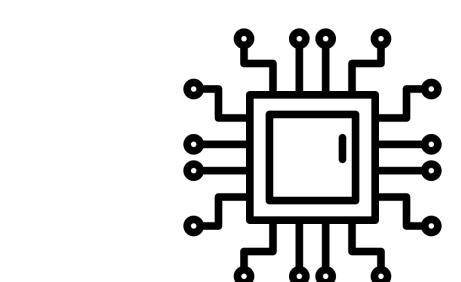
A good AI model can predict correct values even if for inputs / embeddings it has never seen before.

Careful consideration is required:

- Only unseen data should be used for evaluation
- Evaluation should be done multiple times
- Data used for evaluation should be representative of data in the real world
- The metric used for evaluation must capture underlying properties of the data
- AI decisions are not always easily interpretable

Representation

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$



Embedding

Weighing

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$



Parameter Learning

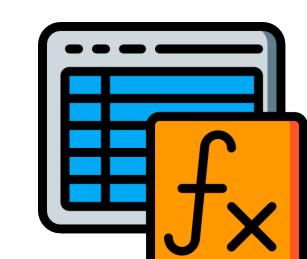
Output

Reality

vs. 1.7%

= 1.8%

Mortgage Interest Rate
(Number)



Regression

Setting a Mortgage Interest Rate - Metrics & Evaluation

Wait! The process is not yet done, even once the model has finished predicting.

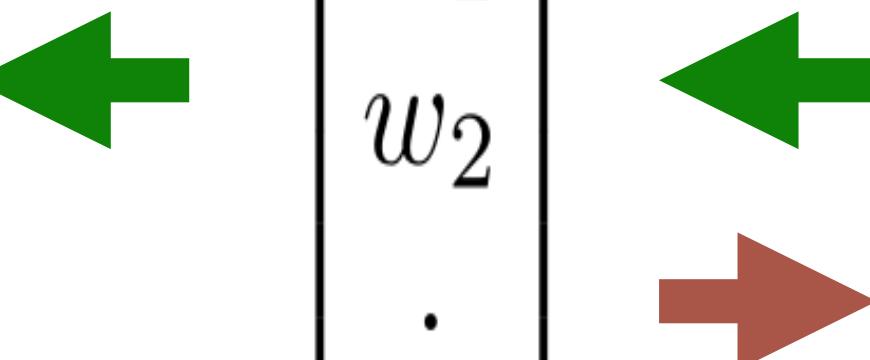
We **criticize** the model to identify further improvements.

Typical model criticism questions:

- Do these predictions make common sense?
- Was all the data used informative?
- Were any important factors missed?
- Do the learnt representations capture properties of underlying data?
- Is the sample size large enough?
- Is the algorithm explainable?
- Is there a tradeoff between prediction performance and explainability of decisions?

Weighing

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

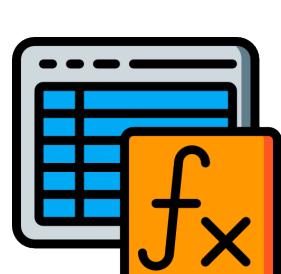


Parameter Learning

Output

$$= 1.8\%$$

Mortgage Interest Rate
(Number)



Regression

Reality

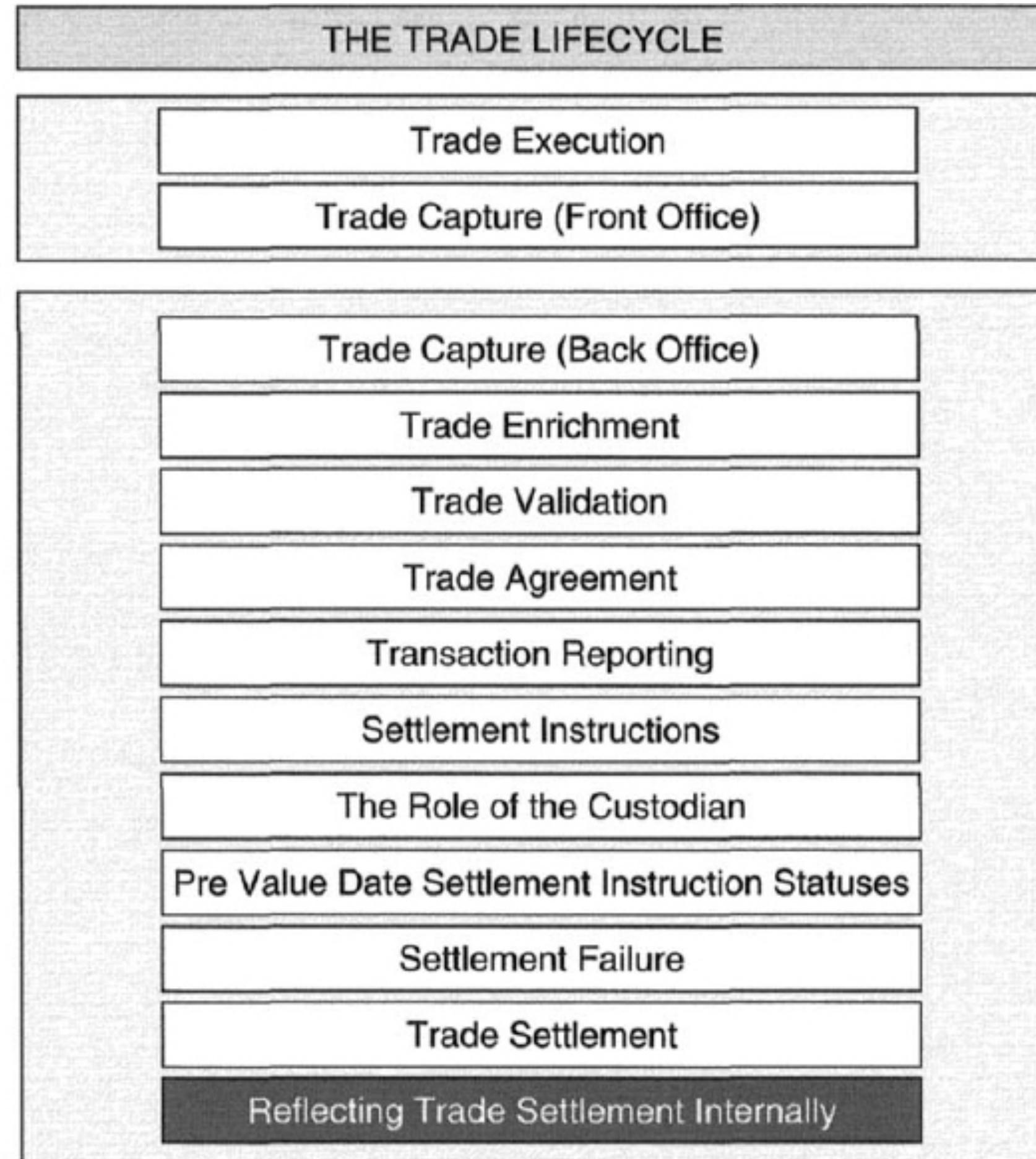
vs. 1.7%

Case Study #1: Trading

- Trade Execution
- Trade Validation
- Know Your Customer



Trading Case Study: Lifecycle of a Trade



A **trader** initiates a trade with the purchase or sale of an asset, commodity, option. The buyer and seller enter into an **agreement**, with a set price, volume, etc.

Operational teams then specify the **logistics** of the trade between parties.

- We compare **traditional** and **AI-driven** approaches to the trading lifecycle.

Key Questions to Keep in Mind

Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

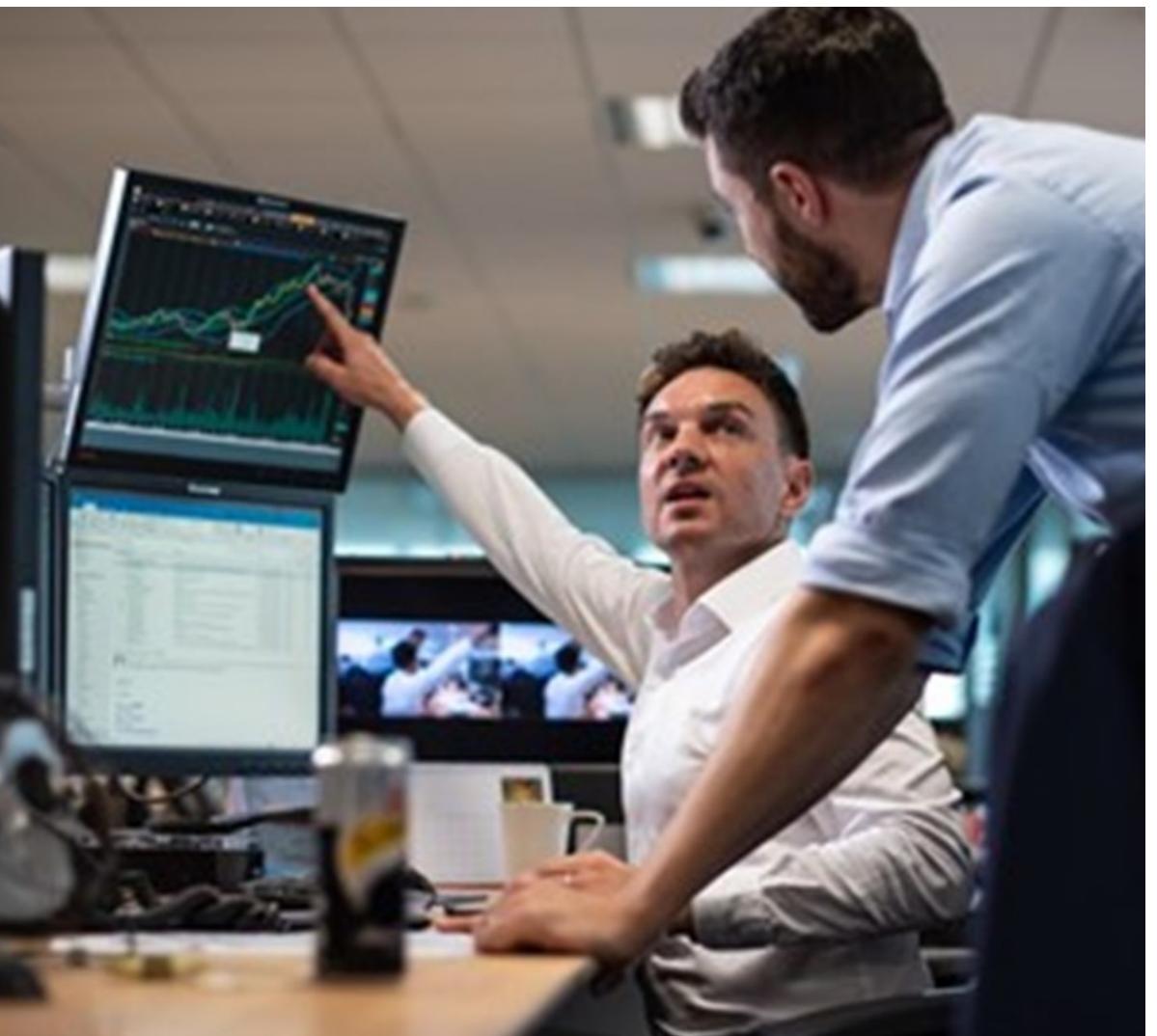
Which parts of the traditional process would benefit from being **data-driven**?

Trading Case Study: Lifecycle of a Trade - Trade Execution

The **trader** is a **human expert** in anticipating price or volatility fluctuations for a given market.

They have experience, know the market and get pricing assistance from quants.

They use their **knowledge** and **intuition** to time the trade and get the best possible price.



Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

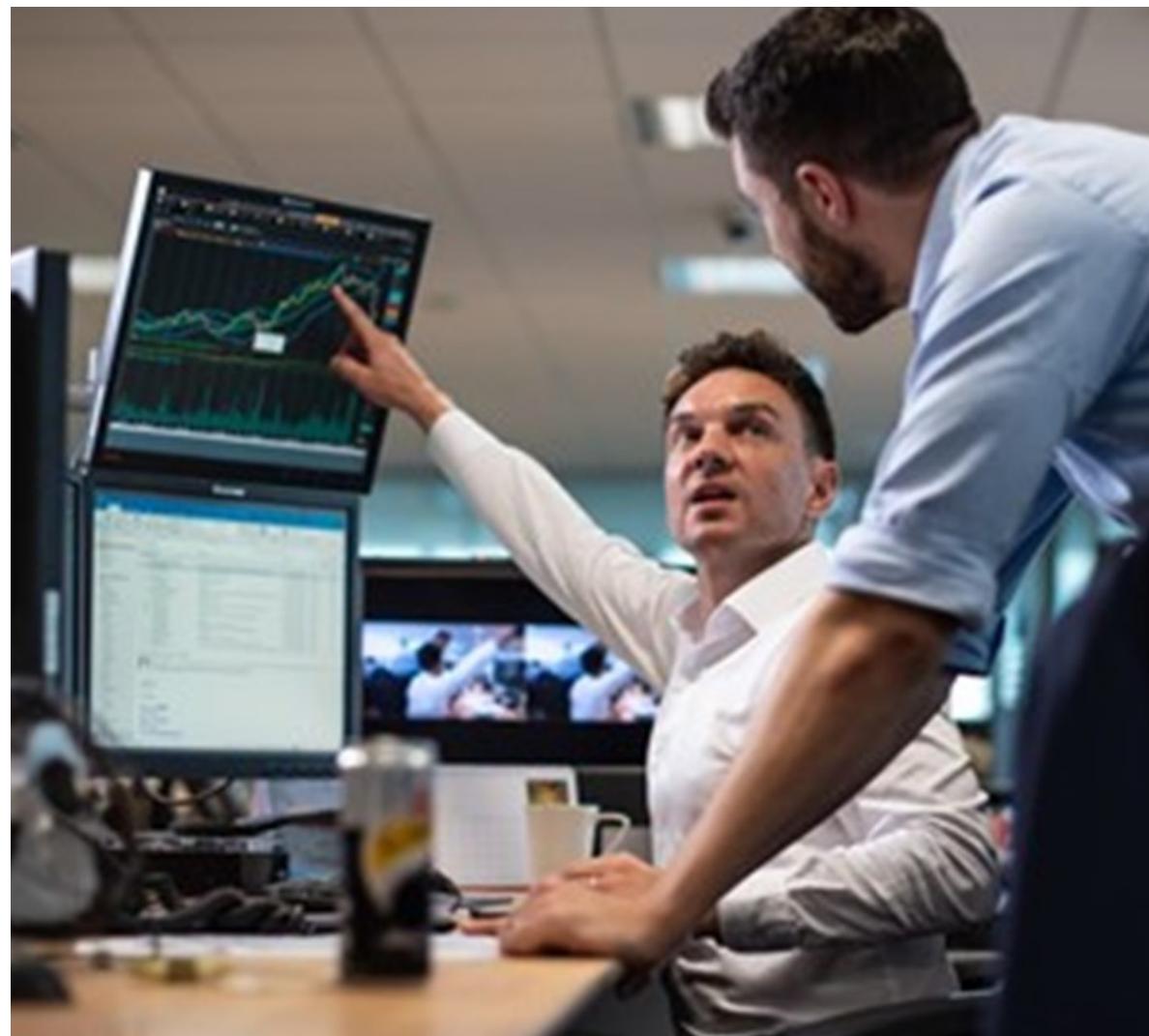
Which parts of the traditional process would benefit from being **data-driven**?

Trading Case Study: Lifecycle of a Trade - Trade Execution

The **trader** is a **human expert** in anticipating price or volatility fluctuations for a given market.

They have experience, know the market and get pricing assistance from quants.

They use their **knowledge** and **intuition** to time the trade and get the best possible price.



Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Bottlenecks (non-exhaustive)

Attention, Speed, Risk Limits, Latency, Emotion, ...

How can AI help?

- **News Aggregation Methods:** Real-time Big Data summarization of the web (Bloomberg, Twitter, Reddit, Government Data).
- **Machine Learning-Based Pricing:** Combine Quant probabilistic pricing methods with Machine Learning-based estimators.
- **Identifying Trading Opportunities:** Automatically compare market patterns to historically similar trading charts.

Trading Case Study: Lifecycle of a Trade - Trade Validation

Operational teams **reconcile** **information** from the trade with the counterpart. Buyer and seller send each other documentation with the agreed upon details.

Ops. teams **verify** that both parties have the exact same information. In some cases, they have two days to **validate the details** and **disclose** the trade to government entities.



Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Trading Case Study: Lifecycle of a Trade - Trade Validation

Operational teams **reconcile information** from the trade with the counterpart. Buyer and seller send each other documentation with the agreed upon details.

Ops. teams **verify** that both parties have the exact same information. In some cases, they have two days to **validate the details** and **disclose** the trade to government entities.



Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Bottlenecks (non-exhaustive)

Time, Complexity, Documentation Errors, ...

How can AI help?

- **Automated Trade Detail Extraction:** Document Understanding algorithms extract prices, amounts, and dates from PDF & images.
- **Document Management:** Classification algorithms automatically categorize document types (invoice, contract, bill of lading, ...).
- **Anomaly Detection:** Algorithms can automatically flag documents with missing or incorrect information.

Trading Case Study: Lifecycle of a Trade - Know Your Customer

Every trade transiting through a Bank is **screened**. Brokers found guilty of approving transactions involving **sanctioned entities** can be fined.

Banks must collect identifiable details such as personal information. These identifiers are crossed with **public watchlists** and **flagged** if they involve sanctioned individuals.



This is a key topic in **crypto**.

Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Trading Case Study: Lifecycle of a Trade - Know Your Customer

Every trade transiting through a Bank is **screened**. Brokers found guilty of approving transactions involving **sanctioned entities** can be fined.

Banks must collect identifiable details such as personal information. These identifiers are crossed with **public watchlists** and **flagged** if they involve sanctioned individuals.

This is a key topic in **crypto**.

Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?



Bottlenecks (non-exhaustive)

Data Complexity, Volume, Public Data, Regulatory Changes

How can AI help?

- **Entity Recognition:** Matching algorithms can disambiguate companies and persons appearing under multiple names.
- **Public Watchlist Monitoring:** Automatic aggregation of data from public sources can help quickly adapt to geopolitical events.
- **Anomaly Detection:** Algorithms reduce operational workload by automatically reviewing and labelling transactions flagged by algorithms.

Case Study #2: Investment Banking

- Client Onboarding
- Due Diligence
- Portfolio / Loan Management



Investment Banking Case Study: Banking Lifecycle

FRONT OFFICE	
Investment Banking Division (IBD)	
Product Groups	Industry Groups
Mergers and Acquisitions (M&A)	Healthcare
Financing / Global Capital Markets	Real Estate
Debt Capital Market	Transportation
Equity Capital Market	Utility and Energy
Leveraged Finance	Natural Resources
Capital Market	Industrials

MIDDLE OFFICE	
Risk Management	
Treasury	
Others	

Corporate and Investment Banks are **banks for companies**.

Companies raise money to grow, either organically or through acquisitions. Banks provide companies with **debt**, assist them in **raising equity**, or **advise** on potential deals.

- We compare **traditional** and **AI-driven** approaches to the banking lifecycle.

Key Questions to Keep in Mind

Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Investment Banking Case Study: Client Onboarding

The **investment banker** works on behalf of a company looking to sell part or all of its business. They cover companies in a specific sector and are experts in **matching** companies from this sector with buyers.

In the deal preparation phase, the banker estimates a **value** for the deal, **notifies** potential buyers, and keeps only the best matches.



Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Investment Banking Case Study: Client Onboarding

The **investment banker** works on behalf of a company looking to sell part or all of its business. They cover companies in a specific sector and are experts in **matching** companies from this sector with buyers.

In the deal preparation phase, the banker estimates a **value** for the deal, **notifies** potential buyers, and keeps only the best matches.



Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Bottlenecks (non-exhaustive)

Speed, High Demand, Market Appetite, Marketing, ...

How can AI help?

- **Recommendation Systems:** Given historical data of deals involving similar buyers and sellers, what is the likelihood of this deal going through?
- **News Aggregation & Sentiment Analysis:** Which companies are likely to have an appetite for deals based on information they shared publicly?
- **Anomaly Detection:** Which potential buyers can be instantly rejected based on informative metrics?

Investment Banking Case Study: Due Diligence

Once the **banker** has identified a suitable buyer, they grant them exclusive access to the client's **financial** and **operational** information. All relevant information is quickly and thoroughly checked to minimize deal uncertainty.

The banker and buyer draft a legal contract for the purchase and agree on a deal **valuation**.



	1.20	0.34	3.25	4.8	8.59	1.7	0.02
56	8.25	3.25	4.8	3	6.05	10.25	
3	10	25.6	12.59	17.98	15.26	129.85	0
18.44	20.77	5.86	3.96	6.6	1		
3	1.5	4			0.5	11	
0	0.5	0	0.37	0	0	0	
2.7	53.32	2.36	0.3	1.21	2.7	22.06	
9964.9	9964.76	11060	13945.79	14851.18	17625.9	19138.99	
149.99	211.18	54.31	453.65	229.93	59.97	139.96	
	Apr	May	Jun	Jul	Aug	Sep	Oct
13359.77	14016.76	1694.89	12901.21	12625.01	13686.73	213.05	
925.61	1232.46	1046.6	1152.52	1210.19	2180.86	2100	
2990.29	3408.59	445.21	3400	2956.12	3779.39	325.32	
340.83	445.02	491.75	442.9	443.92	603	74.39	
8953.85	8323.28	228.76	5744.81	4654.11	6468.39	6983.6	
1675.65	1859.25	178.12	1914.77	1830.85	2268.69	165.45	
911.7	860.27	13.35	979.59	847.94	1067.62	1163.01	
482.46	561	51.83	515.79	558.06	645.75	549	
419.47	390.96	39.32	403.78	402.73	329.75	367.56	
57.72	80.6	42	87.88	35.36	74	85.28	
1.24	0.99		17.86	1.88	37	1.3	
1	0.75		0.25	3.70	2.5	0	
196.66	313.82	14	51	710.8	794.06		
173.81	308	22.03	191.87	172.88	153.71	119.41	
0.2	14.44	0	20.7	0.19	0		
30.8	16.55	23.4	30.25	28.35	45.7		
20.33	15.4	15.92	29.29	18.99	44.92		
7	7.9	1.26	0.62	1.72	35.5	238.59	

Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Investment Banking Case Study: Due Diligence

Once the **banker** has identified a suitable buyer, they grant them exclusive access to the client's **financial** and **operational** information. All relevant information is quickly and thoroughly checked to minimize deal uncertainty.

The banker and buyer draft a legal contract for the purchase and agree on a deal **valuation**.



	Apr	May	Jun	Jul	Aug	Sep	Oct
13359.77	14016.76	1694.89	12901.21	12625.01	13686.73	213.05	
925.61	1232.46	1046.6	1152.52	1210.19	2180.86	2100	
2990.29	3408.59	445.21	3400	2956.12	3779.39	325.32	
340.83	445.02	491.75	442.9	443.92	603	74.39	
8953.85	8323.28	228.76	5744.81	4654.11	6468.39	6983.6	
1675.65	1859.25	78.12	1914.77	1830.85	2268.69	165.45	
911.7	860.27	3.35	979.59	847.94	1067.62	1163.01	
482.46	561	51.83	515.79	558.06	645.75	549	
419.47	390.96	39.32	403.78	402.73	329.75	367.56	
57.72	80.6	42	87.88	35.36	74	85.28	
1.24	0.99	0	17.86	1.88	37	1.3	
1	0.75	0	0.25	3.70	2.5	0	
196.66	313.82	14	51	51	710.8	794.06	
173.81	308	22.03	191.87	172.88	153.71	119.41	
0.2	14.44	0	20.7	0.19	0		
30.8	16.55	23.4	30.25	28.35	45.7		
20.33	15.4	15.92	29.29	18.99	44.92		
7	7.9	1.26	0.62	1.72	35.5	238.59	

Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Bottlenecks (non-exhaustive)

Discretion, Modeling Issues, Data Quality, Market Risk, ...

How can AI help?

- **Contract Understanding:** Algorithms can scan contracts to avoid clauses which shift the balance of the deal.
- **Compliance & Anomaly Detection:** Client information may contain accounting issues which affect the deal price.
- **Pricing:** Machine Learning can help generate realistic scenarios to stress-test financial models.

Investment Banking Case Study: Portfolio / Loan Management

Aside from M&A, banks lend money to corporate clients to help them execute their strategy. **Loan monitoring** teams keep track of all the loans the Bank makes and fees it receives.

Multiple banks can participate in a loan, in which case they receive information on all the **other banks'** participation.



Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Investment Banking Case Study: Portfolio / Loan Management

Aside from M&A, banks lend money to corporate clients to help them execute their strategy. **Loan monitoring** teams keep track of all the loans the Bank makes and fees it receives.

Multiple banks can participate in a loan, in which case they receive information on all the **other banks'** participation.



Where are the **bottlenecks** in the traditional process?

What is stopping the Bank from **scaling** its action?

Which parts of the traditional process would benefit from being **data-driven**?

Bottlenecks (non-exhaustive)

Data Quality, Competition, Yield, Attention, ...

How can AI help?

- **Document Understanding:** Loans involving many banks generate large amounts of documentation.
- **Market Intelligence:** Valuable intel on competing banks can be extracted from information companies disclose.
- **Analytics:** Automatically identify low-fee / low-yield loans to reallocate capital to more profitable clients.

Frontiers of AI in Finance

- LegalTech & Contract Understanding
- Document Management & Search Engines
- Causality & Interpretable Machine Learning



LegalTech & Contract Understanding

Agreements between **buyers** and **sellers** are made legally binding with contracts.

- In Banking, large deals or large loans come with equally **long contracts**.
- In Trading, contracts related to **complex products** (options, swaps, derivatives) cause many headaches for Trade Validation and Risk teams.

The contract is perhaps the most important document, but can hide unfair clauses. If a bad deal is signed, it can become valid!

Example Financial use-cases

Does this contract allow me to sell my participation in this loan?

What does this company mean by “Entry into a Material Agreement”?

The screenshot shows a software interface for managing contracts. At the top, there's a toolbar with icons for file operations, search, and navigation. Below the toolbar, the title of the document is displayed: "Asset Management and Consulting Agreement - Remington Employers Management Corp. and Ashford Financial Corp.". The main content area contains the text of the agreement, which is a standard legal document with sections like "ASSET MANAGEMENT AND CONSULTING AGREEMENT", "RECITALS", and "ARTICLE I ENGAGEMENT". To the right of the document text, there's a sidebar titled "Extract" which lists various contract details. These details include:

- ContractType 2: Management Agreement
- CounterParty: Remington Employers Management Corporation, Ashford Financial Corporation
- Effective Date: May 15, 2003
- Signatory: Monty Bennett
- Jurisdiction: DALLAS COUNTY, TEXAS
- Termination: No
- Confidentiality: No
- Indemnity: Yes
- Severability: Yes

Clause extraction software can identify provisions in contracts and summarize them into simple English

Document Management & Search Engines

Advances in Natural Language Processing make it possible to directly **query** documents using free text.

Combined with a **recommendation system**, this can effectively create a search engine across all files in a database.

This could allow users to query information from files without having to open and read the files manually.

Example Financial use-cases

Which files in the database mention our 2018 credit agreement with L'Oréal?

Did this company share any information on competing banks in their quarterly reports?

What differences exist between these two versions of the same file?



Causality & Interpretable Machine Learning

Causal Inference is a framework used in Statistics to **quantify** the **effect** variables have on outcomes.

Given the importance of **audits** in Finance, causality could help bridge the gap between opaque AI models and explainable financial models.

Causal Inference plays a large role in **Economics**, but is challenging to implement in practice and requires extreme care to avoid misinterpreting results.

Pharmaceutical companies use Causal Inference...

...to measure the effect of a treatment (vaccine, antibiotic, etc.) on a patient's medical condition.

Social media companies use Causal Inference...

...to measure the effect of a UI change on user behavior.

Economists use Causal Inference...

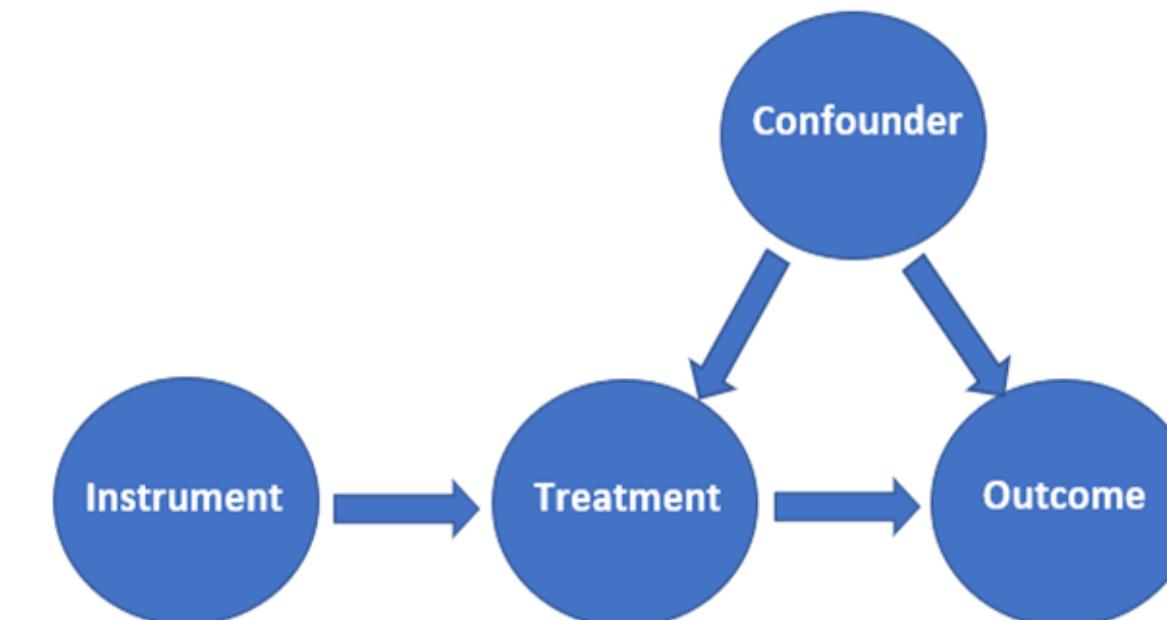
...to evaluate the effect of policies on labor markets.

Example Financial use-cases

What is the effect of a stock-split on the stock's price?

Why was this transaction flagged as fraudulent?

What is the effect of a mortgage rate increase on defaults?



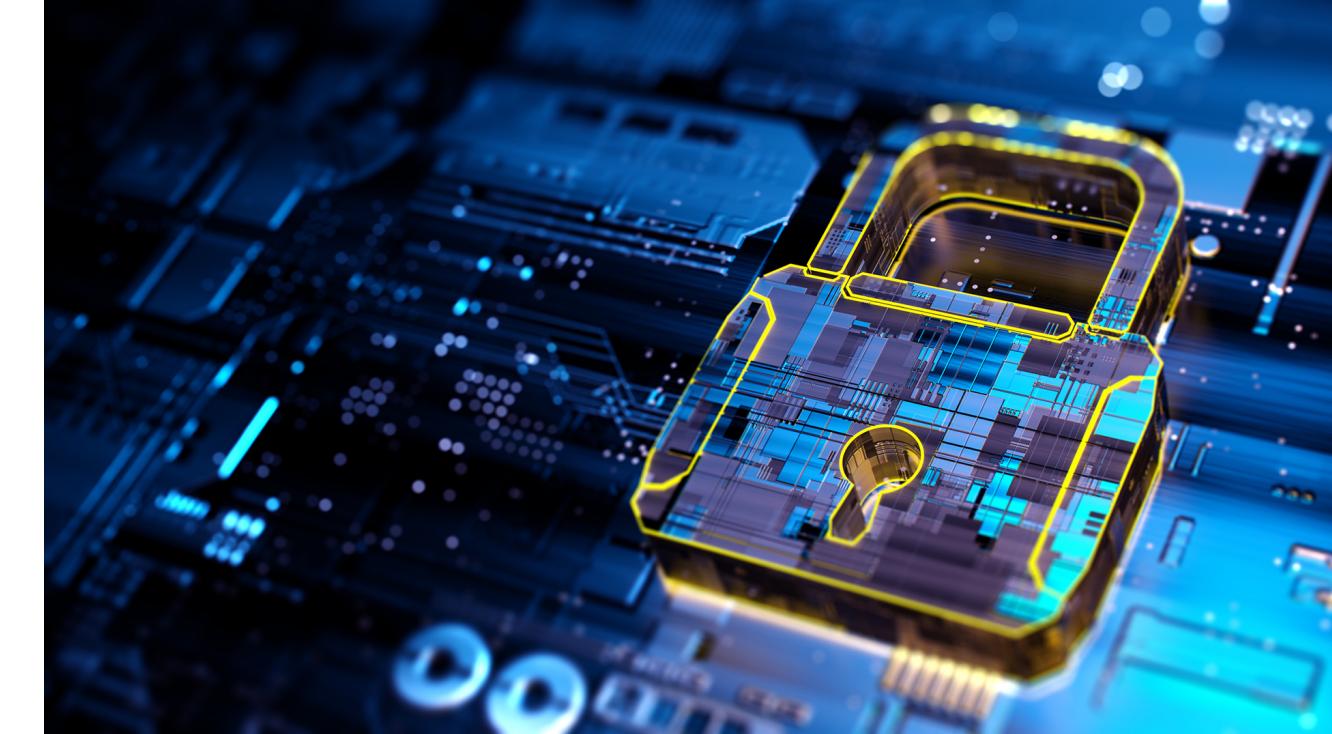
Risks for AI Adoption in Finance



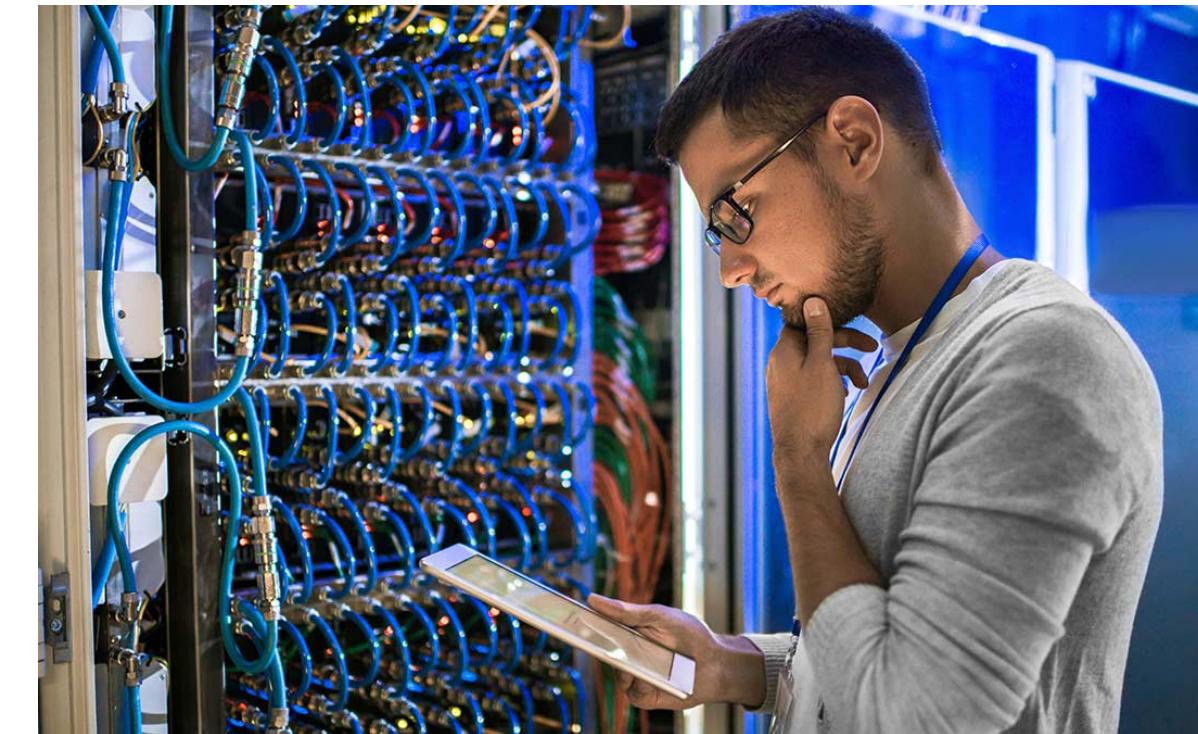
Risks From AI in Finance



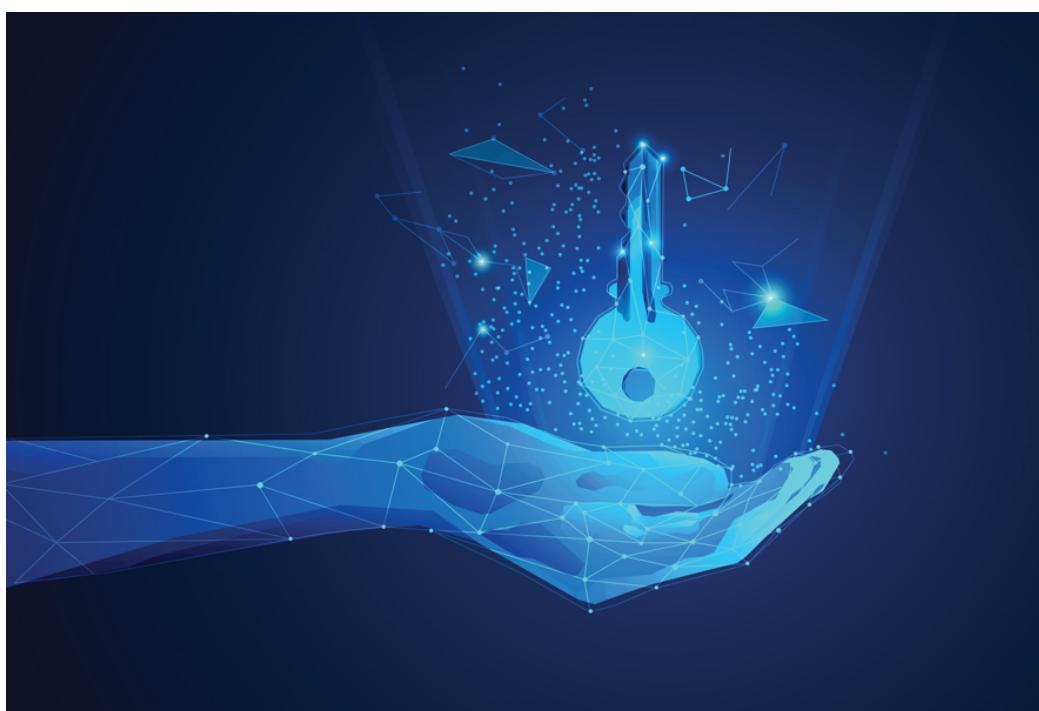
Audit & Explainability



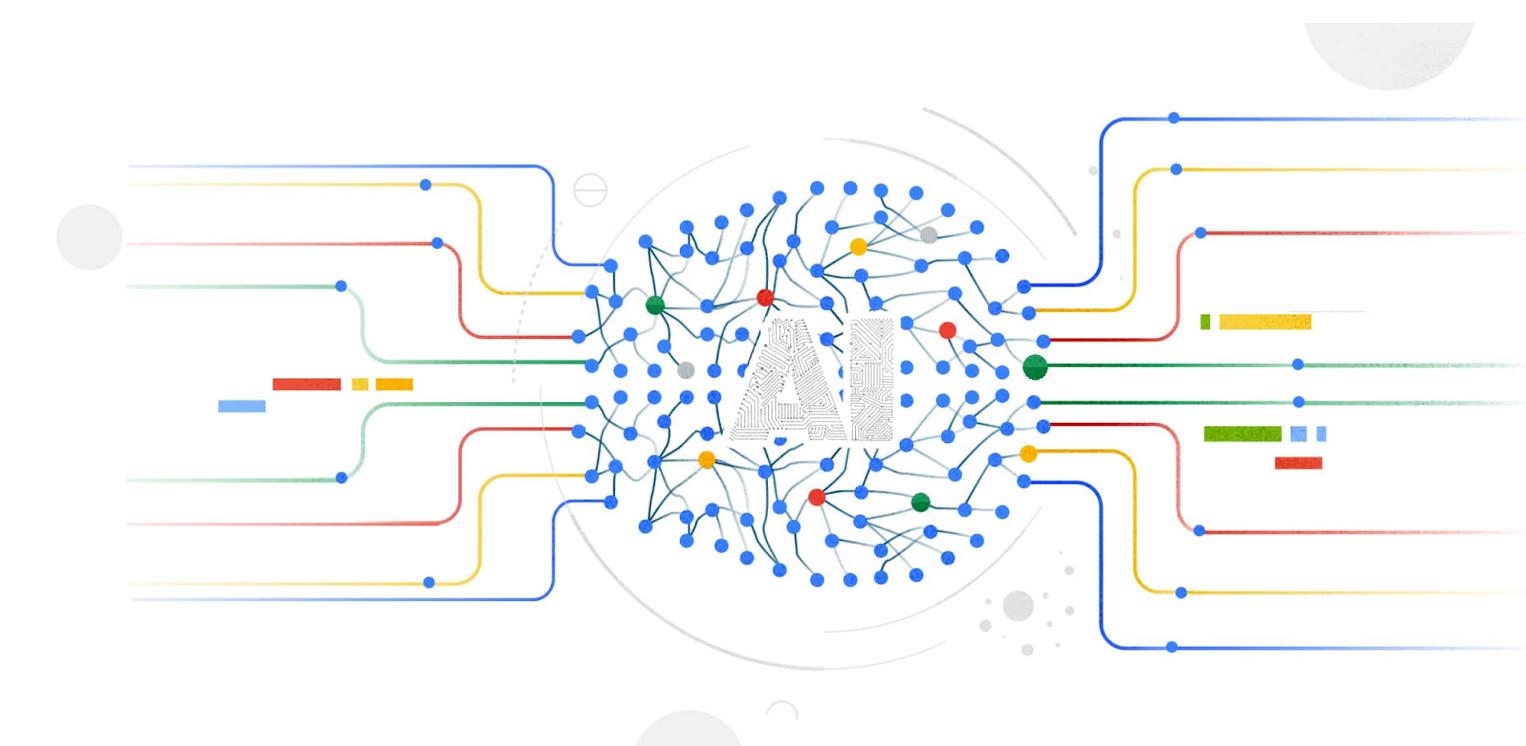
Cybersecurity



Infrastructure



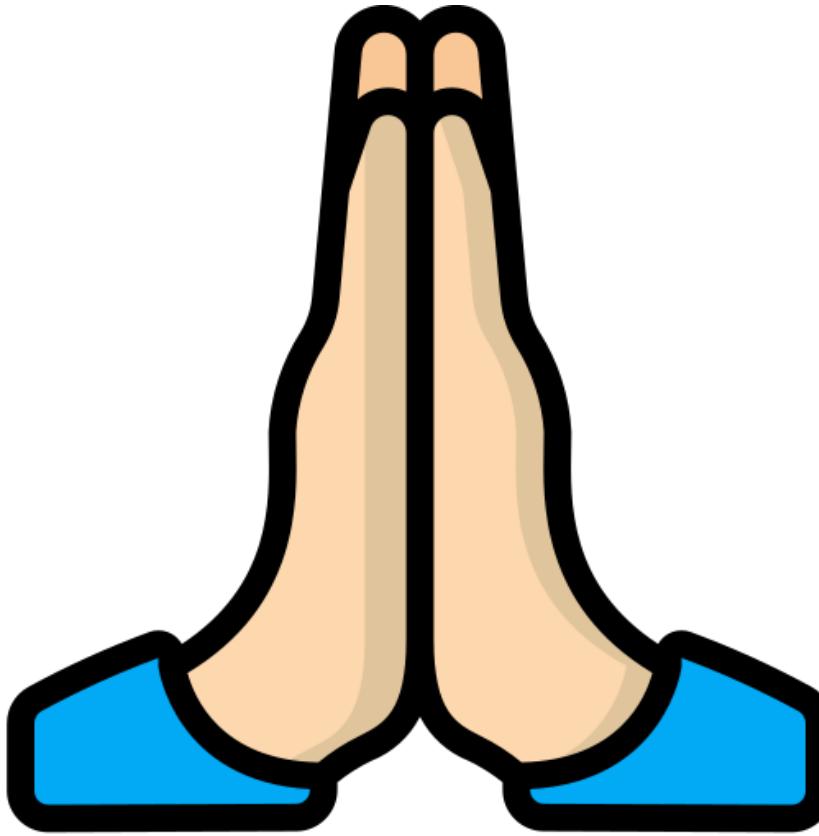
Data Privacy



Model Errors

Final Thoughts

- Many AI in Finance topics we didn't cover (Asset Management, Pricing, Insurance, Retail Banking, Research, etc.)
- All feedback welcome (positive and negative)
- Slides available at github.com/mtchibozo/hec-international-fintech
- Many of these challenges will be up to you to solve, start coding!



Thank You!