AI in Finance: From Prediction to Production

Data Science with Python – BSc Course (Supplementary Lecture)

Data Science Program

BSc Course

90 Minutes

Can Machines Read the Market?

Both are sometimes right. Only one can explain why.

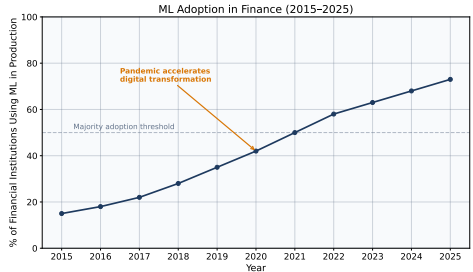
This is the central tension of AI in finance: raw performance versus the ability to explain a decision.

Why Is AI Suddenly Everywhere in Finance?

Three forces, all measurable

- Financial data volume grew ~10x in the last decade (tick data, filings, alternative data)
- Cloud compute cost per GPU-hour has fallen roughly 70% since 2020
- Algorithmic trading is now ~75% of US equity volume (TABB Group, 2024)
- Regulators now *require* model risk management (SR 11-7 in the US, EU AI Act in Europe)

Read the chart: Notice the steep acceleration after 2020. The pandemic forced many institutions to digitise overnight, which turned into permanent ML adoption.



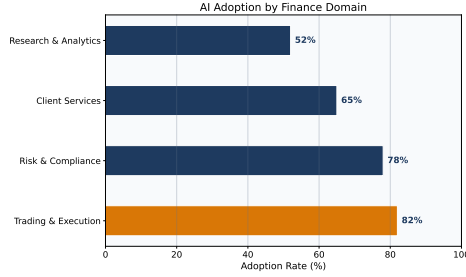
Source: composite of McKinsey "State of AI" and BIS 2024 AI chapter. "In production" means deployed in live decision systems, not R&D experiments.

What Does the AI Finance Ecosystem Look Like?

Four domains, four maturities

- **Trading & Execution:** algorithmic strategies, optimal execution, market making
- **Risk & Compliance:** credit risk, market risk, AML (anti-money laundering), fraud
- **Client Services:** robo-advisory, chatbots, personalised recommendations
- **Research & Analytics:** sentiment analysis, earnings forecasting, ESG scoring

Read the chart: Risk & Compliance leads because regulators demand it. Client Services is growing fastest thanks to LLMs.



These four domains structure the rest of this lecture. By slide 38 you will have seen at least one real example in each.

What Will You Be Able to Do After 90 Minutes?

Five concrete abilities

1. **Identify** the four domains where AI is deployed in finance and name one real system (e.g., JPMorgan LOXM, BlackRock Aladdin) in each
2. **Explain** how logistic regression, gradient boosting, and neural networks are used in credit scoring and fraud detection, including their trade-offs
3. **Evaluate** claims about AI trading performance using backtesting bias, overfitting, and market efficiency as critical tools
4. **Distinguish** between explainable and black-box models, and articulate why regulators care about the difference
5. **Design** a high-level ML pipeline for a financial prediction task, identifying where each course topic (L01-L48) applies

---

These five abilities are your self-assessment checklist. If you can do all five by slide 45, you have succeeded.

What Do You Already Know That Applies Here?

Course Module	Key Lessons	Finance AI Application
Python + Data (L01-L12)	DataFrames, NumPy, time series	Tick-data pipelines, OHLCV (Open-High-Low-Close-Volume) processing
Statistics (L13-L16)	Distributions, hypothesis testing, correlation	Risk modelling, factor analysis, signal testing
Visualisation (L17-L20)	Matplotlib, Seaborn, storytelling	Dashboards, model monitoring, client reports
Regression (L21-L24)	Linear, regularisation, metrics	Factor models, pricing, VaR (Value at Risk)
Classification (L25-L28)	Logistic, trees, imbalance	Credit scoring, fraud detection, churn
Unsupervised (L29-L32)	Clustering, PCA, anomalies	Customer segmentation, anomaly detection
Deep Learning (L33-L36)	Neural nets, CNNs, NLP	Sentiment analysis, document processing
Deployment (L41-L48)	APIs, monitoring, ethics	MLOps in production, bias auditing

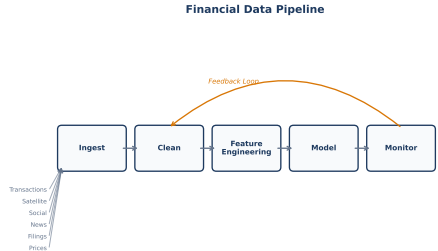
You are not starting from zero. Every row in this table is a tool you already have. This lecture shows you where to use each one.

Where Does Financial ML Data Actually Come From?

Six input streams, one pipeline

- Six sources: market data, fundamentals, alternative data (satellite, web), news/text, transactions, filings
- JPMorgan runs 300+ AI use cases across billions of data points daily (2024 annual report)
- Data is noisy, non-stationary, and has survivorship bias (failed firms vanish)
- Pipeline: *Ingest -¿ Clean -¿ Features -¿ Model -¿ Monitor*

Read the chart: Notice the feedback arrow from Monitor back to Clean – models degrade, retraining is continuous.



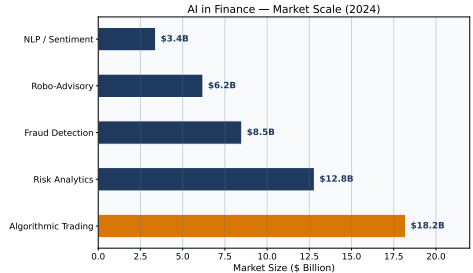
Folklore: "80% of ML engineering is data work." In finance this applies even more strongly because data quality directly drives profit and loss.

How Big Is AI in Finance Today?

Trillion-dollar scale, already

- Algorithmic trading: ~75% of US equity volume (TABB Group, 2024)
- Robo-advisors: ~$2.7 trillion AUM (Assets Under Management), ~90 million users globally
- BlackRock Aladdin: risk analytics for ~$21 trillion across 200+ clients
- AI fraud prevention saves an estimated $10–15 billion per year (Nilson Report 2024)

Read the chart: Even the "smallest" bar (fraud prevention) represents billions. This is not niche technology.



Figures as of 2024. Most categories are growing at 15–25% annually. Aladdin's figure counts assets analysed, not assets managed by AI alone.

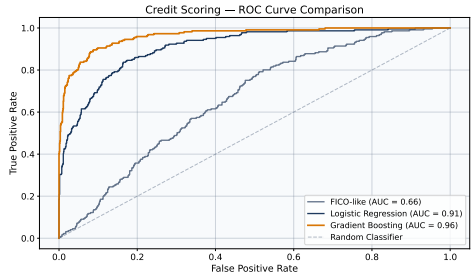
Part 2: Prediction

From Regression to Deep Learning

Can a Model Decide Who Gets a Loan?

From FICO to logistic to gradient boosting

- FICO: weighted sum of 5 factors (payment history 35%, amounts owed 30%, length 15%, new credit 10%, mix 10%)
- ML: logistic regression on 50–200 features, outputs default probability p; decline if $p > 0.5$
- Gradient boosting adds 5–10 AUC points over FICO (Upstart, Fed working papers)
- Higher AUC (Area Under the Curve) is not the same as fair or explainable



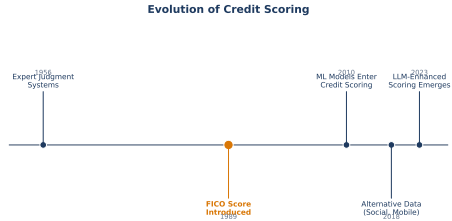
Read the chart: Higher and further left = better. Amber (gradient boosting) sits above slate (FICO) across the whole range.

AUC (Area Under the ROC Curve): 0.50 = random, 1.00 = perfect. AUC of 0.85 means a defaulter outranks a non-defaulter 85% of the time.

Why Did Banks Replace FICO with Machine Learning?

Dimension	FICO	ML-Based
Features	5 predefined	50–200+
Model	Linear sum	Boosting / NN
Explainability	Transparent	Needs SHAP
Performance	AUC 0.70–0.75	AUC 0.78–0.85
Bias risk	Zip-code proxy	New proxies

Read the chart: It took 40 years from expert judgement to FICO, but only 10 years from FICO to LLM-enhanced scoring.



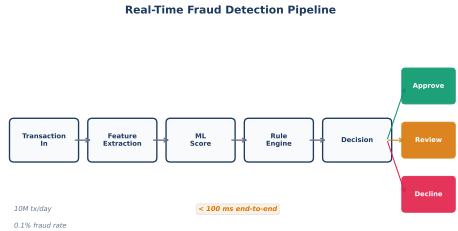
Most US banks still run FICO and ML in parallel, using ML to flag borderline cases the traditional model might miss.

How Do Banks Catch Fraud in 100 Milliseconds?

Real-time classification, extreme imbalance

- Global card fraud ~$33B per year; AI prevents $10–15B more (Nilson Report 2024)
- Class imbalance: only ~0.1% of transactions are fraudulent (1 in 1,000)
- Decision in <100 ms per transaction
- Models: random forest and gradient boosting dominate; deep sequence models for complex patterns

Read the chart: The score is only *one* input. A rule engine combines it with business rules before the final decision.

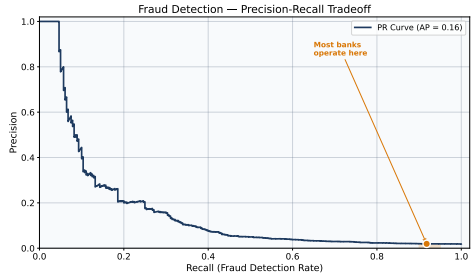


The 100 ms latency budget limits deep-learning use. Model size and serving speed matter more than marginal accuracy gains.

What Happens When Fraud Detection Gets It Wrong?

Two very different costs

- False positive (FP): legitimate transaction blocked. Cost: customer frustration, lost revenue
- False negative (FN): fraud approved. Cost: financial loss, reputational harm
- Typical bank target: recall >90%, willing to accept precision of 5–15%
- Optimal threshold depends on the *cost ratio*: if FN costs 100x FP, you accept more false alarms



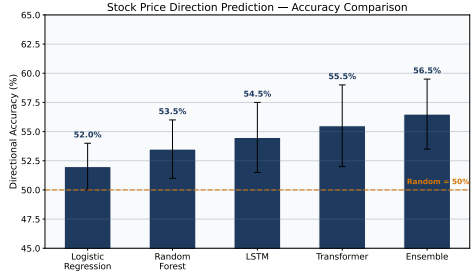
Read the chart: The shaded zone is where most banks operate. Moving left (higher recall) costs precision very quickly.

A model with 95% recall and 10% precision means ~90% of flagged transactions are not fraud. This is why your card sometimes gets blocked when travelling.

Is Stock Prediction the Holy Grail of AI Finance?

Realistic numbers, not headlines

- Best published directional accuracy: ~52–58% for daily predictions (Gu, Kelly, Xiu 2020, *Review of Financial Studies*)
- 55% can still be profitable *if* transaction costs are low and volume is high
- Common models: LSTM (Long Short-Term Memory networks), Transformers, ensembles
- Out-of-sample: most published results do not survive realistic transaction costs



Read the chart: The dashed line at 50% is random guessing. Every bar is only slightly above it, and the error bars are WIDE.

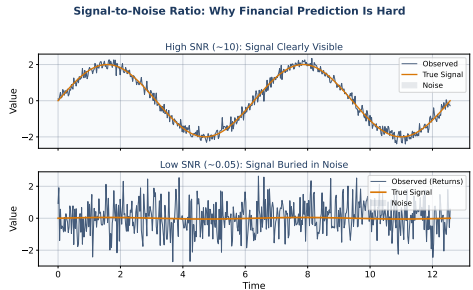
The gap between 50% and 55% looks small, but consistent 55% directional accuracy with disciplined risk management can be very valuable. The problem is consistency.

Why Is Stock Prediction Fundamentally Harder Than Image Recognition?

Three structural reasons

- *Efficient Market Hypothesis (EMH):* predictable patterns get arbitraged away quickly
- *Non-stationarity:* bull / bear / crisis regimes have different statistics
- *Signal-to-noise ratio (SNR):* stock returns ˜0.05 vs images ˜10
- Reflexivity: every prediction you act on *changes* the market

Read the chart: Top panel: a clean pattern jumps out. Bottom panel: same signal amplitude, now buried in noise.



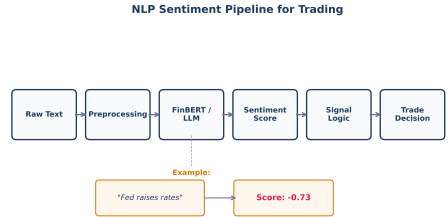
SNR (Signal-to-Noise Ratio) is the most under-appreciated concept in financial ML. A model cannot learn what is not there.

How Do You Turn a News Article into a Trading Signal?

Text in, score out, decision next

- Text sources: news, earnings call transcripts, social media, SEC filings, analyst reports
- Pipeline: *Raw Text -¿ Preprocessing -¿ FinBERT or LLM -¿ Score -¿ Signal Logic -¿ Trade*
- Evidence: sentiment has *modest* predictive power for intraday to weekly horizons
- FinBERT (Araci 2019) is the standard academic baseline, pre-trained on financial text

Read the chart: Notice the gap between "score" and "trade". A negative sentiment does not automatically mean "sell" – signal logic is a separate, often rule-based, layer.



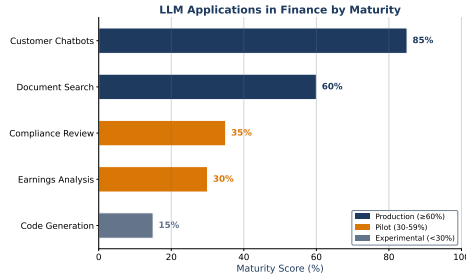
FinBERT beats general BERT on financial sentiment by ~10–15 F1 points, because domain pre-training teaches it that "hawkish" and "dovish" are not emotions.

What Can LLMs Do in Finance Besides Chat?

Four real use cases, all back office

- **Document search:** Morgan Stanley's GPT-4 advisor indexes 100,000+ research reports
- **Compliance review:** automated contract review, detecting policy violations
- **Earnings analysis:** summarising calls, flagging language shifts quarter over quarter
- **Code generation:** writing and reviewing trading strategy code (with human oversight)

Read the chart: LLMs entered finance via low-risk chatbots and are migrating, slowly, toward higher-stakes decisions as maturity grows.

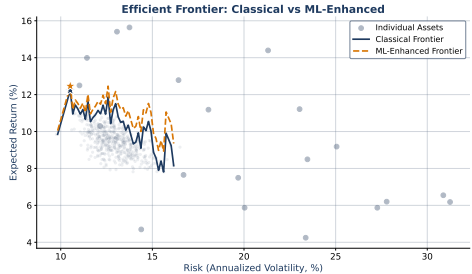


Risks: hallucination (LLMs confidently invent financial numbers), prompt injection, and data leakage. BloombergGPT (Wu et al., 2023) is the first large-scale financial LLM published in detail.

Can Machine Learning Beat Classical Portfolio Theory?

Small improvements, big impact

- Markowitz (1952): maximise $\mu^\top w - \frac{\lambda}{2} w^\top \Sigma w$ (μ = returns, Σ = covariance)
- ML 1: shrinkage covariance (L22 regularisation) beats sample covariance
- ML 2: ML-predicted returns replace historical averages
- ML 3: reinforcement learning for dynamic rebalancing (no closed form)



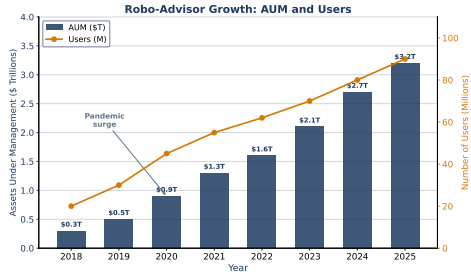
Read the chart: The amber (ML-enhanced) frontier sits slightly above the navy (classical) one. Better inputs, not magic.

Underappreciated fact: the practical gain from ML in portfolio optimisation comes mostly from better covariance estimation, not from better return prediction.

Can an Algorithm Replace Your Financial Advisor?

Cheap, accessible, rule-driven

- Robo-advisors automate: risk profiling, asset allocation, tax-loss harvesting, rebalancing
- Typical fee: 0.25–0.50% vs 0.75–1.25% for traditional advisors
- Global AUM ~$2.7 trillion (2024), projected ~$6 trillion by 2027
- Performance: roughly matches benchmark indices after fees; advantage is *access*



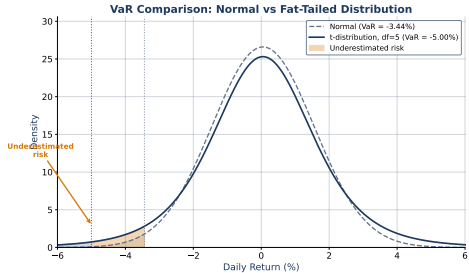
Read the chart: Steepest growth is 2020–2022. AUM growth outpaces user growth after 2022 – average account size is rising.

Robo-advisors are not replacing human advisors. They are expanding the market to clients who could not previously afford advisory fees at all.

Why Do Simple VaR Models Fail in a Crisis?

VaR (Value at Risk) in plain English

- Definition: the maximum loss you expect with 99% confidence over one day
- Formula (normal): $\text{VaR}_{99\%} = -(\mu - z_{0.99} \cdot \sigma)$ where μ = expected return, σ = volatility
- **Example:** \$10M portfolio, $\mu = 0$, $\sigma = 1.5\%$ daily $\Rightarrow \text{VaR}_{99\%} = \$10\text{M} \times 2.33 \times 1.5\% \approx \$349{,}500$
- Real returns have fat tails (L14). Normal VaR *underestimates* extreme losses.



Read the chart: The dashed (normal) distribution vs solid (actual, fat-tailed). The gap at the left tail is the under-estimation.

The 2008 financial crisis was partly caused by VaR models that assumed normally distributed returns.

Neural-network VaR can capture fat tails, but introduces new risks: complexity and overfitting.

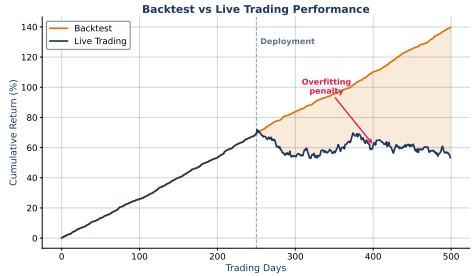
Part 3: Challenges

When AI Meets Reality

Why Do Backtesting Stars Fail in Live Trading?

The cardinal sin of quant finance

- Backtesting = testing a strategy on historical data; easy to *memorise* patterns that will not repeat
- A backtest that looks too perfect is *suspicious*: it is probably fitting noise
- In live trading, future data has patterns the model has never seen
- Rule of thumb: expect live performance to be 30–50% worse than backtest



Read the chart: Everything left of the dashed deployment line is hindsight. Everything right is reality. The gap is the "overfitting penalty".

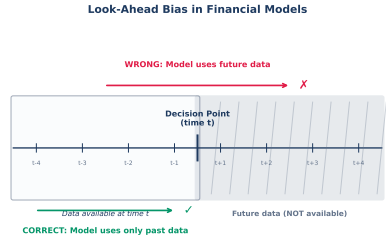
The best protection is walk-forward validation: train on data before time t, test on data after t, and never peek.

How Can a Model "Cheat" Without You Noticing?

Look-ahead bias, defined

- Using information in training that would not have been available at prediction time
- Obvious example: using tomorrow's earnings to predict today's price
- Subtle example: using the *revised* GDP number (released months later) instead of the initial estimate
- Prevention: strict temporal ordering, point-in-time data, walk-forward validation

Read the chart: The rose arrow crosses the "now" line – that is the forbidden move. The green arrow stays in bounds.

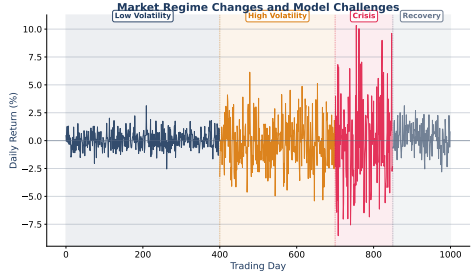


Look-ahead bias can inflate backtest returns by 10–50%. It is the most common error in academic financial ML papers.

What Happens When Markets Change Their Rules?

i.i.d. does not hold in finance

- i.i.d. (independent and identically distributed): training data and future data come from the same distribution
- In finance this is *false*: bull, bear, crisis, recovery regimes all have different statistics
- Correlations break: stocks that move together in calm markets diverge in crises
- Consequence: a model trained 2015–2019 can fail completely in March 2020



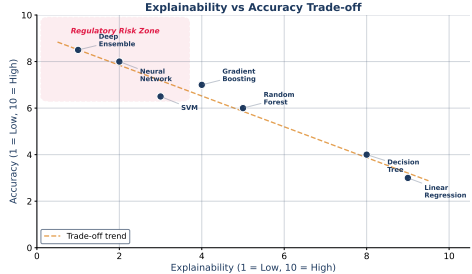
Read the chart: Notice how the spread (volatility) is dramatically wider in the rose (crisis) band. A model trained only on the navy band would be shocked.

Regime boundaries are only clear in hindsight. Detecting a regime shift in real time is itself a hard prediction problem.

Why Does Complexity Cost Trust in Financial Models?

The accuracy-explainability trade-off

- More complex models are often more accurate *and* harder to explain
- A bank that declines a loan must tell the applicant *why* (US ECOA adverse action notice; EU GDPR right to explanation)
- EU AI Act classifies credit scoring as "High Risk", requiring transparency
- XAI (Explainable AI): techniques like SHAP and LIME that open the box *after the fact*



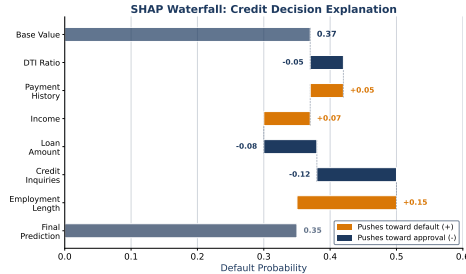
Read the chart: The upward-left trend shows the trade-off. The key question: where on this curve does regulation *let* you operate?

SHAP and LIME partially explain complex models, but "partially" may not satisfy a regulator. In finance, the highest-accuracy model is not always the deployable one.

How Do You Explain a Model's Decision to a Regulator?

SHAP in plain English

- SHAP (SHapley Additive exPlanations): each feature gets a contribution to the prediction (game theory)
- Base value ϕ_0 + feature contributions ϕ_j = final prediction
- **Example:** base 0.35; DTI 42% +0.15; income -0.08; credit 7yr -0.05 $\Rightarrow$ 0.37 declined
- SHAP explains *one* prediction, not the whole model



Read the chart: Largest bar (DTI +0.15) is the main reason for denial. Drop DTI to 30% and this applicant is approved.

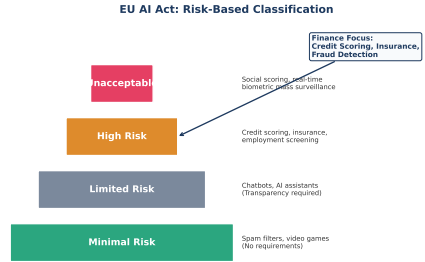
SHAP is mathematically additive: base + all contributions = prediction exactly. Regulators prefer SHAP over other XAI methods because of this guarantee.

How Does the EU AI Act Reshape Financial AI?

Risk tiers, not bans

- EU AI Act (Regulation 2024/1689), phased rollout 2024–2027
- Tiers: Unacceptable (banned), High Risk (strict), Limited (transparency), Minimal (free use)
- Finance: credit scoring, insurance pricing, fraud = High Risk
- Penalty: up to 7% of global revenue or EUR 35M (whichever is greater)

Read the chart: Finance sits in the "High Risk" band – not banned, but heavily regulated.



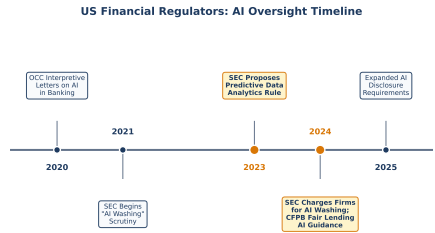
High Risk = conformity assessment + docs + human oversight + transparency + monitoring. Compliance cost for a large bank is estimated at EUR 5–15 million.

How Do the SEC and CFPB Regulate AI Differently?

Enforcement first, legislation later

- US: no comprehensive AI law (as of 2025). Sector-specific rules and enforcement instead.
- SEC: "AI washing" enforcement; Predictive Data Analytics proposal (2023)
- CFPB (Consumer Financial Protection Bureau): fair lending rules apply to AI
- FINRA Notice 24-09: broker-dealers must supervise AI like human recommendations

Read the chart: Timeline accelerates sharply after 2023 – four agencies regulating different aspects in parallel.

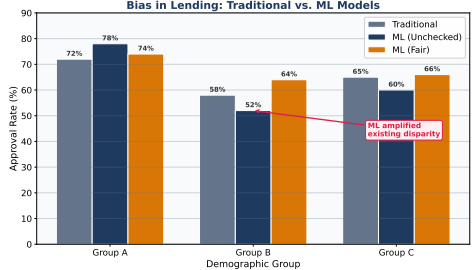


The "enforcement first" US approach means regulatory uncertainty: rules evolve while banks build. Global institutions must satisfy both EU and US regimes.

Can a Model Discriminate Without Seeing Race?

Proxies make blind models biased

- Proxy discrimination: zip code correlates with race, purchase history with gender
- Historical bias: if past lending was discriminatory, training data encodes it
- Disparate impact: equal accuracy, *different* approval rates is still unfair
- Test before deploy: compare approval rates, error rates, and SHAP distributions across demographic groups



Read the chart: The navy bars (ML without constraints) show *wider* gaps than the slate bars (traditional). ML *amplified* the historical bias.

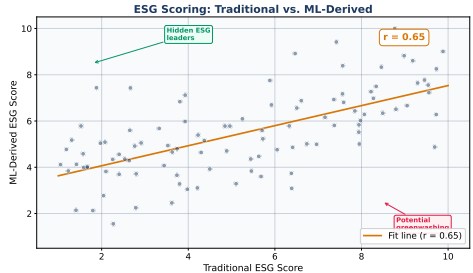
Removing race from the features does not prevent discrimination. The model finds proxies. Fairness requires active testing and fairness constraints, not just passive feature exclusion.

Can Machine Learning Measure Climate and ESG Risk?

NLP, satellites, and a data-quality problem

- ESG scoring: NLP on corporate reports, news, filings for E (Environmental), S (Social), G (Governance) signals
- Climate risk: ML for transition risk (policy) and physical risk (extreme weather)
- Satellite + ML: monitor deforestation, emissions, construction in near real time
- Challenge: ESG data is noisy; providers disagree; greenwashing corrupts labels

Read the chart: Correlation ˜0.6–0.7 means ML and human raters mostly agree – but outliers matter.



Different providers give the same company wildly different ESG scores. MSCI, Sustainalytics, and S&P Trucost can disagree by 40+ points.

Part 4: From Lab to Production

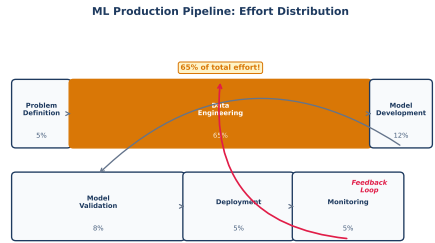
Pipelines, Monitoring, and Case Studies

What Are the Six Stages of a Production ML Pipeline?

Six stages, not equal in effort

- **1. Problem Definition** (5%): what are we predicting, and why?
- **2. Data Engineering** (65%): ingest, clean, feature engineering
- **3-4. Model Dev + Validation** (18%): train, test, fairness, stress
- **5-6. Deployment + Monitoring** (12%): serving, drift, retraining triggers

Read the chart: Box width is proportional to effort. The Data Engineering box dominates.



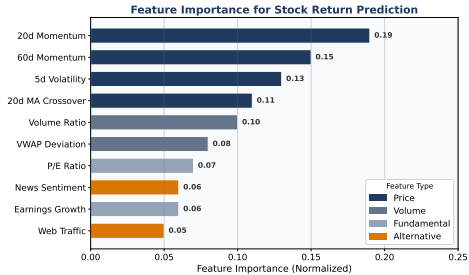
In academia, the interesting part is Stage 3. In industry, the hard part is everything else. A bank may spend 6 months on data engineering for 2 weeks of modelling.

Which Features Matter Most in Financial Prediction?

Five families of features

- **Price:** returns, moving averages, volatility, momentum
- **Volume:** volume ratio, VWAP (volume-weighted avg price) deviation
- **Fundamental:** P/E ratio, book-to-market, earnings growth
- **Alternative / cross-sectional:** sentiment, web traffic, satellite, sector rank

Read the chart: Price-based features typically dominate. Alternative data is useful but rarely the top contributor – despite the hype.

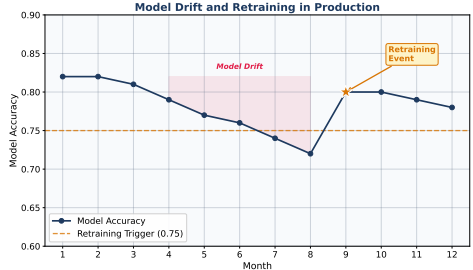


Feature engineering is where domain expertise compounds. A finance-trained data scientist will build better features than a general ML engineer – because they know what drives asset prices.

When Does a Deployed Model Stop Working?

Three kinds of drift

- **Data drift:** input feature distributions shift (e.g., interest rates move from 0% to 5%)
- **Concept drift:** the *relationship* between features and target changes
- **Prediction drift:** model outputs shift (e.g., approval rates move)
- Monitor weekly: accuracy, feature distributions, output distributions; set retraining triggers



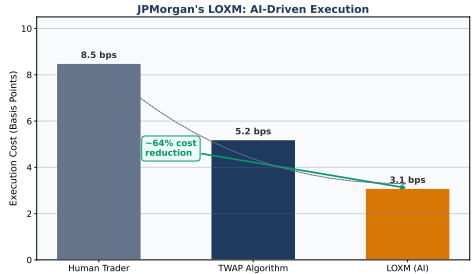
Read the chart: Model accuracy starts at 0.82 and gradually drops. When it crosses the dashed alert threshold, the team retrains (star marker).

A model that is 82% accurate at deploy and 68% six months later is not "broken" – it is normal. What would be broken is not detecting the drop.

What Does JPMorgan's LOXM Actually Do?

Reinforcement learning in real markets

- LOXM: JPMorgan's AI for optimal trade execution (large orders without moving the price)
- How: a reinforcement-learning agent learns to split orders into smaller pieces and time them to minimise impact
- Result: better average prices than human traders *and* the standard TWAP (Time-Weighted Average Price) benchmark
- Deployed across equities, in production since 2017



Read the chart: LOXM (amber) achieves the lowest cost, beating both human traders and the TWAP benchmark.

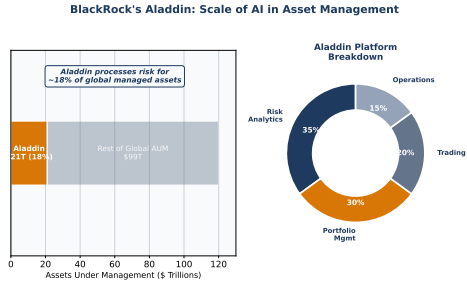
Basis points (bps) look small: 1 bp = 0.01%. At JPMorgan's volume, even 1 bp saved is tens of millions per year.

Why Is BlackRock's Aladdin Systemically Important?

One platform, a lot of AUM

- Aladdin = Asset, Liability, Debt, and Derivative Investment Network
- Scale: risk analytics for ~$21 trillion across 200+ institutional clients (BlackRock, 2024)
- ML components: risk factor modelling, scenario analysis, NLP on filings, optimisation
- Systemic risk concern: if one platform manages risk for ~17% of global managed assets, a model failure could ripple market-wide

Read the chart: The amber slice is not BlackRock's own assets – it is the share of the *global* managed asset pool that runs on Aladdin's models.



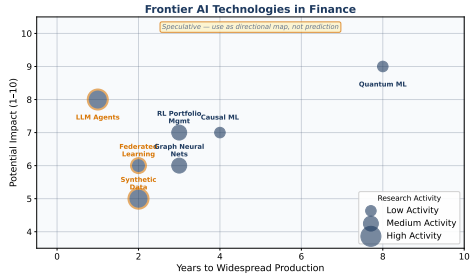
Aladdin is both the most successful ML platform in finance and the best argument for why AI concentration risk matters. Success and systemic risk are two sides of the same coin.

Where Is AI Finance Heading in the Next Five Years?

Four frontiers, four maturities

- **LLM-native finance:** autonomous agents that read, reason, and act on documents (pilot stage)
- **RL at scale:** beyond execution – portfolio management, market making, dynamic hedging (research)
- **Federated learning:** train across institutions without sharing customer data (pilot)
- **Quantum-ML optimisation:** portfolio and risk simulation (research, 5–10 years out)

Read the chart: Bubble size = current research activity. LLMs are closest to production; quantum is furthest.



Prediction is hard, especially about the future. This chart is speculative. Use it as a directional map, not as a fact sheet.

Where Can Data Science Skills Take You in Finance?

Four careers, one common recipe (technical + domain)

- **Quantitative Analyst ("Quant"):** build predictive models for trading, risk, pricing. Skills: L13–L16 (statistics), L21–L24 (regression), L12 (time series). Employers: hedge funds, banks, prop trading firms
- **ML Engineer (Finance):** build and maintain ML pipelines in production. Skills: L01–L06 (Python), L07–L12 (data), L41–L44 (deployment). Employers: banks, fintechs, asset managers
- **Risk Analyst:** credit, market, and operational risk models. Skills: L14 (distributions), L25–L28 (classification), VaR (slide 20). Employers: banks, insurers, regulators
- **FinTech Data Scientist:** consumer-facing AI (robo-advisors, fraud, credit). Skills: full stack L01–L48. Employers: startups, neobanks, payment companies

---

All four paths need both technical skills (what you learned) and financial domain knowledge. The intersection is where the value lives.

Part 5: Review and Synthesis

Formulas, Myths, and Mastery

Which Eight Formulas Should You Remember?

Prediction

1. Logistic Regression: $p = 1/(1 + e^{-(\beta_0+\beta_1 x_1+\cdots)})$
Probability of default. Threshold at 0.50.

2. Sharpe Ratio: $\mathrm{SR} = (R_p - R_f)/\sigma_p$
Risk-adjusted return. R_f = risk-free rate.

3. Mean-Variance: $\max_w \mu^\top w - \frac{\lambda}{2} w^\top \Sigma w$
Markowitz: balance return (μ) and risk (Σ).

4. Value at Risk: $\mathrm{VaR}_\alpha = -\mathrm{Quantile}_\alpha(R)$
Max expected loss at confidence α (e.g., 99%).

Evaluation & Explainability

5. Precision: $\mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$
Of flagged transactions, how many were actually fraud?

6. Recall: $\mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$
Of actual fraud cases, how many did we catch?

7. SHAP Additivity: $f(x) = \phi_0 + \sum_{j=1}^{M} \phi_j$
Prediction = base value + sum of feature contributions.

8. AUC: $P(\mathrm{score}(y{=}1) > \mathrm{score}(y{=}0))$
Probability a random defaulter outranks a non-defaulter.

Do not memorise these. Know when to use each and how to interpret its output. Every formula appeared earlier with a worked example.

Which Four Myths About AI in Finance Should You Retire?

Myth 1: "AI will replace all traders"
AI handles execution and pattern detection. Humans still set strategy, manage risk, and handle novel situations. Goldman Sachs still employs thousands in trading – with different tasks than in 2000.

Myth 2: "More data always means better models"
Non-stationary data from the wrong regime is *worse* than no data. A model trained 2010–2019 can learn patterns that reversed in 2020. Relevance beats quantity.

Myth 3: "Deep learning beats everything"
For tabular financial data, gradient boosting (XGBoost, LightGBM) often matches or beats deep learning. DL's real advantage is unstructured data: text, images, sequences.

Myth 4: "Backtests prove a strategy works"
Backtests prove a strategy *would have worked* on historical data. Overfitting and regime change typically cost 30–50% of the backtest edge live.

Misconceptions 5–8 are on the next slide. Quant interviewers use these myths to filter candidates.

Which Four More Myths Should You Retire?

Myth 5: "If the model is accurate, it is fair"
A model can be 90% accurate overall but deny loans to one demographic group at twice the rate of another. Accuracy and fairness are independent dimensions – both must be measured.

Myth 6: "Removing race/gender from features eliminates bias"
Proxies (zip code, spending patterns, school) carry the same information. The model learns the proxies. Active fairness testing is required.

Myth 7: "AI in finance is new"
Algorithmic trading started in the 1970s. Neural nets for credit scoring date to the 1990s. What is new is *scale* (cheap compute) and *scope* (LLMs, alternative data).

Myth 8: "Regulators are against AI"

Regulators want AI that is transparent and fair, not banned. The EU AI Act explicitly *allows* finance AI under "High Risk". The goal is trustworthy AI, not no AI.

Test yourself: can you explain why each myth is wrong using specific evidence from this lecture?

Can You Answer These Four Questions Without Looking Back?

Q1: A bank's credit scoring model has AUC = 0.85 but approves Group A at 80% and Group B at 55%. Is this model "good"?

Can You Answer These Four Questions Without Looking Back?

Q1: A bank's credit scoring model has AUC = 0.85 but approves Group A at 80% and Group B at 55%. Is this model "good"?

- **A1:** Accurate but potentially *unfair*. High AUC does not guarantee equal treatment. The 25-point gap in approval rates requires a disparate-impact investigation.

Can You Answer These Four Questions Without Looking Back?

Q1: A bank's credit scoring model has AUC = 0.85 but approves Group A at 80% and Group B at 55%. Is this model "good"?

- **A1:** Accurate but potentially *unfair*. High AUC does not guarantee equal treatment. The 25-point gap in approval rates requires a disparate-impact investigation.

Q2: Your stock model achieves 62% accuracy in backtesting but 51% in the first live month. Three possible explanations?

Can You Answer These Four Questions Without Looking Back?

Q1: A bank's credit scoring model has AUC = 0.85 but approves Group A at 80% and Group B at 55%. Is this model "good"?

- **A1:** Accurate but potentially *unfair*. High AUC does not guarantee equal treatment. The 25-point gap in approval rates requires a disparate-impact investigation.

Q2: Your stock model achieves 62% accuracy in backtesting but 51% in the first live month. Three possible explanations?

- **A2:** (1) Overfitting to historical patterns; (2) look-ahead bias in training; (3) regime change between training and live periods.

Can You Answer These Four Questions Without Looking Back?

Q1: A bank's credit scoring model has AUC = 0.85 but approves Group A at 80% and Group B at 55%. Is this model "good"?

- **A1:** Accurate but potentially *unfair*. High AUC does not guarantee equal treatment. The 25-point gap in approval rates requires a disparate-impact investigation.

Q2: Your stock model achieves 62% accuracy in backtesting but 51% in the first live month. Three possible explanations?

- **A2:** (1) Overfitting to historical patterns; (2) look-ahead bias in training; (3) regime change between training and live periods.

Q3: An LLM summarises an earnings call and states "Revenue grew 15%." The actual figure is 12%. What kind of AI risk is this?

Can You Answer These Four Questions Without Looking Back?

Q1: A bank's credit scoring model has AUC = 0.85 but approves Group A at 80% and Group B at 55%. Is this model "good"?

- **A1:** Accurate but potentially *unfair*. High AUC does not guarantee equal treatment. The 25-point gap in approval rates requires a disparate-impact investigation.

Q2: Your stock model achieves 62% accuracy in backtesting but 51% in the first live month. Three possible explanations?

- **A2:** (1) Overfitting to historical patterns; (2) look-ahead bias in training; (3) regime change between training and live periods.

Q3: An LLM summarises an earnings call and states "Revenue grew 15%." The actual figure is 12%. What kind of AI risk is this?

- **A3:** Hallucination. LLMs generate plausible but factually wrong text. Dangerous in finance because wrong numbers drive investment decisions.

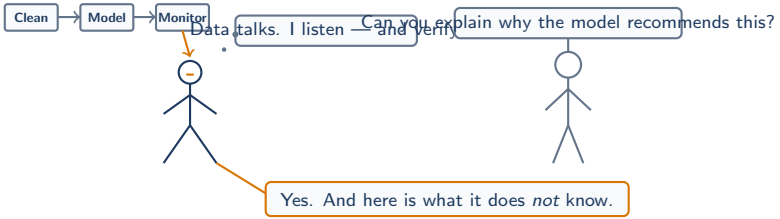
If you can answer all three without flipping back, you have met the learning objectives from slide 5.

What Should You Read Next?

Interest Area	Course Connection	Further Reading
Credit scoring	Revisit L25 (Logistic), L26 (Trees)	Siddiqi, *Credit Risk Scorecards* (Wiley)
Fraud detection	Revisit L27 (Metrics), L28 (Imbalance)	Bolton & Hand, *Statistical Fraud Detection*
Stock prediction	Revisit L21–L24, L33–L36	Gu, Kelly, Xiu (2020), *Rev. Fin. Studies*
NLP in finance	Revisit L37–L40	Araci (2019), *FinBERT*
Explainability	Revisit L47 (Ethics)	Molnar, *Interpretable Machine Learning* (free online)
Risk management	Revisit L14 (Distributions)	Hull, *Risk Management and Financial Institutions*
Careers	All of L01–L48	QuantNet; Financial Data Professional certificate

Every row connects back to your course. AI in finance is not a separate discipline – it is your existing skills applied to the highest-stakes domain in business.

Data Talks. I Listen – and Verify.



The difference between a data scientist and a dangerous data scientist is the word "verify". Never trust a model you cannot explain.