Finance and AI

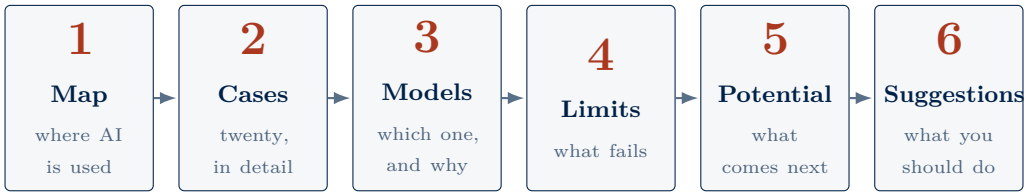
Twenty cases, the models behind them, and what to do next

Fatih Kansoy

Worcester College, University of Oxford

Day 4

Today: six parts



Parts 1 to 4 before the break. Parts 5 and 6 after.

Day 3 gave you four questions. Today they get twenty answers

How much?

prices, cash flow, risk

regression

Will it happen?

default, fraud, distress

classification

Groups or flags?

customers, transactions

unsupervised

What does text say?

filings, news, calls

language models

And one test, applied to every case today:

What is predicted? What evidence supports it? Which errors matter? Who acts on it?

How to read every case in this lecture

Every case gets one slide, and the same five fields.

Task	approve or decline a payment
Type	supervised classification
Model	boosted trees, then a deep network
Data	payments; labels are chargebacks
Result	32% less fraud, on the vendor's figure

By the end you should be able to fill these in for a case you have never seen.

And every number carries its evidence tier.

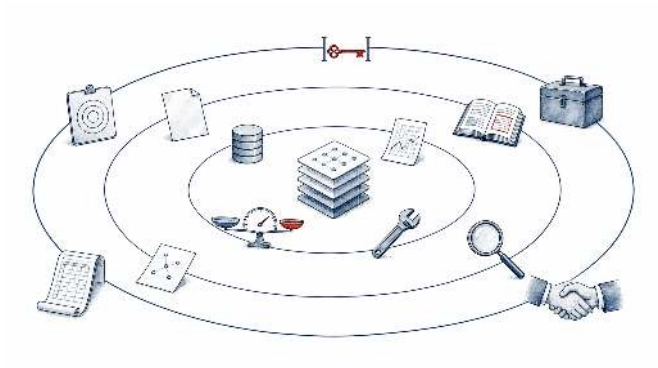
REGULATOR OR COURT	strongest
PEER-REVIEWED	referees checked it
COMPANY SAYS SO	they measured themselves
VENDOR CASE STUDY	they are selling it
PROJECTION	not a measurement

Ask of every number: who measured it, and what would a bad result have looked like?

Part 1

Where AI is used in finance

Eight domains, four kinds of learning, and one chain from data to decision.



Eight domains, and what each one predicts

Domain	What is predicted	Typical model
Credit and lending	will this person repay?	logistic regression, boosting
Fraud and crime	is this payment real?	boosted trees, deep networks
Trading and investing	what is next month's return?	trees, neural networks
Insurance and risk	how large is the claim?	GLM, boosting
Business and customers	how many will we sell?	boosted trees
Documents and banking	what does this text say?	transformers
Economics and policy	what is the economy doing now?	factor models
Geoeconomics and climate	where are the ships and risks?	vision, networks

Eight industries. The same four questions underneath all of them.

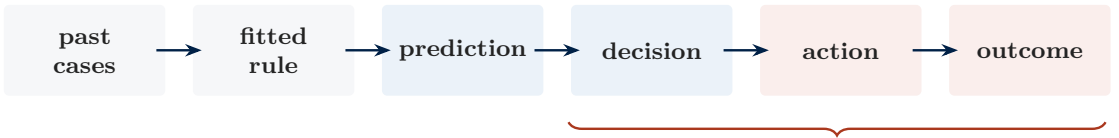
Four kinds of learning, four kinds of finance job

Learning type	What you must have	Finance example
Supervised	past cases with answers	default, fraud, claim size
Unsupervised	no answers at all	customer groups, odd transactions
Self-supervised	raw text or images	language models, satellite images
Reinforcement	a sequence of actions	order execution, hedging

Which one you may use is decided by **what answers you can get**, and how fast they arrive.

The label decides the method, not the other way round.

A model creates value only through a chain



A prediction is not a decision, and a decision is not an outcome.

Part 2

Twenty cases, eight domains

For each one: the task, the learning type, the model, the data and its labels, and the measured result.

Contents: the twenty cases

Credit and lending	1 Credit scoring since 1941 · 2 Digital footprints · 3 Buy now, pay later
Fraud and crime	4 Card fraud in milliseconds · 5 Money-laundering networks · 6 Deepfake payment fraud
Trading and investing	7 Predicting stock returns · 8 Execution and hedging
Insurance and risk	9 Pricing: accuracy against auditability · 10 Telematics
Business and customers	11 Demand forecasting · 12 Platform pricing · 13 Customer support
Documents and banking	14 Contract review · 15 Firmwide assistants · 16 Bespoke or general model
Economics and policy	17 Nowcasting the economy · 18 Text as data
Geoeconomics and climate	19 Ships and sanctions · 20 ESG ratings

Then one slide on the same patterns in Chinese finance.

Case 1. Credit scoring: the oldest case, and still simple

Task	approve a loan, and set its rate
Type	supervised classification
Model	logistic regression, since the 1950s
Data	past borrowers; label arrives years later
Result	an industry, and a duty to give reasons

1941 Durand sorts instalment loans · **1956** Fair and Isaac · **1989** the FICO score

Finance did statistical learning for fifty years before anyone said "AI".

Source: Durand, NBER, 1941. REGULATOR OR COURT

Case 2. Digital footprints score people with no credit file

Task	score a customer with no bureau file
Type	supervised classification
Model	logistic regression
Data	ten signals from the purchase itself
Result	AUC 69.6%, against 68.3% for the bureau

69.6% digital footprint

68.3% full bureau score

Device, operating system, time of day, email provider, and how the customer typed their own address.

New data, old mathematics. The gain came from the inputs, not the model.

Source: Berg, Burg, Gombović and Puri, *RFS*, 2020. PEER-REVIEWED

Case 3. Buy now, pay later: whoever owns the labels owns the moat

Task	decide credit at the checkout, in a second
Type	supervised classification
Model	not disclosed
Data	thin data; label is repayment, weeks later
Result	credit decisions with no bureau file pulled

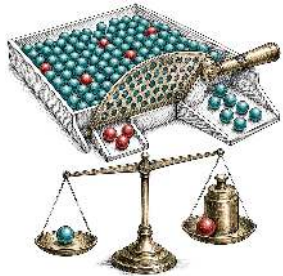
An ordinary lender **reports your repayment** back to the credit bureaus. Some of these lenders do not.

If nobody else holds your outcome data, nobody else can train your model.

Source: Reporting practice varies by firm and year. COMPANY SAYS SO

Case 4. Card fraud: a decision in milliseconds, with no human

Task	approve or decline before the terminal answers
Type	supervised classification
Model	rules, then boosted trees, then a deep network
Data	70tn data points; labels are chargebacks
Result	32% less fraud, on the vendor's own figure



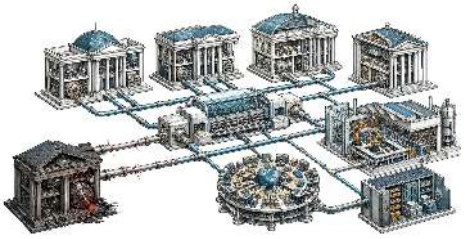
Fraud is perhaps one payment in a thousand. A model approving **everything** is 99.9% accurate, and worthless.

Under heavy imbalance the false positive is the priced object, not accuracy.

Source: Stripe product documentation. VENDOR CASE STUDY

Case 5. Money laundering: find the network, not the payment

Task	flag accounts moving criminal money
Type	unsupervised, then human review
Model	entity resolution and network analysis
Data	payments and ownership records; **no labels**
Result	alerts for investigators, not decisions



The largest fine ever in this field was **not a model failure**. One bank left **92%** of its transaction volume unmonitored, some **$18.3 trillion** over six years, and paid about **$3.09 billion**.

Better AI would not have helped. Nothing was watching in the first place.

Source: US Department of Justice and three other regulators, 2024. REGULATOR OR COURT

Case 6. Deepfake fraud: the attacker has the same tools

Task	impersonate executives on a video call
Type	generative
Model	video and voice generation
Data	public images and recordings
Result	about HK$200 million paid out in one day

An employee joined a call with his “chief financial officer” and several colleagues. **Every other person was generated.**

Identity confirmation is now a payments control, not an IT detail.

Source: Hong Kong Police; the firm confirmed the incident, 2024. PRESS

Case 7. Predicting stock returns: a tiny edge, repeated

Task	next month's return, for every US stock
Type	supervised regression
Model	trees and neural networks vs linear models
Data	30,000 stocks, 1957–2016, 900+ signals
Result	monthly out-of-sample R^2 of 0.33–0.40%

Strategy	Sharpe
Linear benchmark	0.61
Neural network	**1.35**

Long-short decile spread, out of sample, before trading costs.

The model explains about **four thousandths** of monthly variation. It is wrong nearly all the time.

Barely predictable and highly profitable are compatible, across thousands of positions.

Source: Gu, Kelly and Xiu, *RFS*, 2020. PEER-REVIEWED

Case 8. Execution and hedging: the case with no published numbers

Task	split a large order; hedge a book over time
Type	reinforcement learning
Model	a policy learned in a market simulator
Data	order books and simulated prices
Result	**no disclosed effect size, in nine years**

A natural fit: a sequence of actions, each one changing the next. Banks have described these systems since 2017.

Absence of a number is information: usually secrecy, occasionally nothing to report.

Source: Bank announcements and trade press only. COMPANY SAYS SO

Case 9. Insurance pricing: accuracy against auditability

Task	price a motor policy
Type	supervised
Model	boosted trees, against a GLM
Data	1.2 million real quotes
Result	boosting wins, by 0.017 of AUC

Model	Log-loss	AUC
XGBoost	**0.0890**	**0.9064**
Random forest	0.0923	0.8955
Deep learning	0.0936	0.8925
GLM	0.0940	0.8896

The last row is the one a regulator can **read line by line** in a rate filing.

In regulated pricing, explainability is often the binding constraint, not accuracy.

Source: Spedicato, Dutang and Petrini, *Variance* 12(1). PEER-REVIEWED

Case 10. Telematics: safer drivers, or safer driving?

Task	price motor cover from driving behaviour
Type	supervised
Model	not disclosed by the insurer
Data	sensor data from enrolled drivers
Result	lower losses, from **two different causes**

Behaviour	Change
Speeding	−11 to −13%
Hard braking	−16 to −21%
Fast acceleration	−16 to −25%
Phone in hand	no change

Against a randomised control group.

Safer drivers **volunteer** (selection). Or the same drivers **improve** (causation). Only a trial separates the two.

A rate filing cannot tell you which. Somebody had to run the experiment.

Source: University of Pennsylvania trial, *Accident Analysis and Prevention*, 2025. PEER-REVIEWED

Case 11. Demand forecasting: the cleanest evidence in the pack

Task	daily unit sales, 28 days ahead
Type	supervised regression
Model	gradient-boosted trees (LightGBM)
Data	42,840 real Walmart series, 2011–2016
Result	22.4% better than the best benchmark

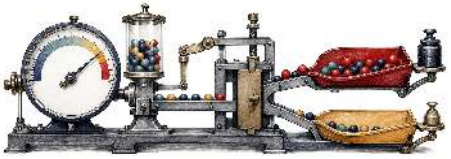
Everyone had the **same data**, the same horizon, and a scoring rule announced in advance. All top-50 entries beat the benchmark.

A public competition on real data beats any vendor's accuracy claim.

Source: M5 competition, *International Journal of Forecasting*, 2022. PEER-REVIEWED

Case 12. Platform pricing: efficiency created, and captured

Task	set a price for each trip, in real time
Type	not disclosed
Model	not disclosed
Data	trip, demand and supply data
Result	median platform share rose from 25% to 29%



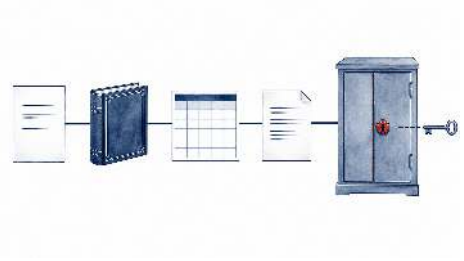
Better matching of supply and demand is a **real** efficiency gain. It is a separate question who keeps it.

Ask two things of a pricing model: is the pie bigger, and who holds the knife?

Source: Study presented at ACM FAccT, 2025. PEER-REVIEWED

Case 13. Customer support: the first properly measured gain

Task	suggest replies to support agents, live
Type	generative language model
Model	a large language model, as an assistant
Data	past conversations and their resolutions
Result	**+15%** productivity, across 5,172 agents



The average hides the finding. In the earlier working paper, gains were concentrated among the **least experienced** staff.

Where these tools help, they help beginners most. In this room, that is everyone.

Source: Brynjolfsson, Li and Raymond, *QJE*, 2025. PEER-REVIEWED

Case 14. Contract review: the first document win

Task	find and extract clauses in loan agreements
Type	supervised
Model	clause extraction, later transformers
Data	contracts; labels marked up by lawyers
Result	a claimed 360,000 lawyer-hours a year

That figure is the bank's own estimate, with **no baseline published**. The academic benchmark has **no human comparison at all**.

The labels here are slow and expensive, which is why this took until 2016.

Source: Firm statements COMPANY SAYS SO ; CUAD benchmark PEER-REVIEWED .

Case 15. Firmwide assistants: adoption is not impact

What the banks disclose	What they do not
over 98% of adviser teams use the assistant	whether the advice improved
it indexes over 100,000 research documents	whether clients were better served
one internal tool reached 200,000 staff	any audited productivity study

This is structural, not dishonest. **Nobody logs the memo that was never written.**

Usage tells you a system is popular. It does not tell you it is valuable.

Source: Firm and vendor disclosures, 2024–2026. COMPANY SAYS SO

Case 16. Bespoke or general? An expensive natural experiment

Task	answer financial questions from text
Type	self-supervised, then fine-tuned
Model	a 50bn-parameter model, trained in-house
Data	a proprietary financial archive
Result	beaten within a year by a general model

One financial question-answering benchmark. The bespoke model did keep an edge on narrower tasks.

Owning the data is not a moat when general capability improves faster.

Source: Comparison reported at EMNLP, 2023. PEER-REVIEWED

Case 17. Nowcasting: measuring the economy before the data arrive

Task	estimate this quarter's growth, weekly
Type	unsupervised dimension reduction
Model	factor model, **not machine learning**
Data	about 30 indicators, arriving at odd times
Result	suspended for two years when COVID broke it

Official data arrive weeks late. A factor model reads many fast indicators and extracts **one common signal**.

It assumes past relationships hold. When they break, it breaks when you need it.

Source: Federal Reserve Bank of New York, suspended 2021, relaunched 2023. COMPANY SAYS SO

Case 18. Text as data: much of it is not machine learning

Method	What it does	Where it is used
Counting keywords	counts words in newspapers	geopolitical risk indices
Finance word lists	fixes generic sentiment tools	filings and earnings calls
Topic models	finds themes without labels	central-bank transcripts
Transformers	reads meaning in context	everything since 2019

"Liability", "tax" and "cost" are **not bad news** in a filing. They are the filing. Generic sentiment tools get this wrong.

Reach for the simplest method that answers the question. A word count can be audited.

Source: Caldara and Iacoviello; Loughran and McDonald; Hansen, McMahon and Prat, *QJE*, 2018. PEER-REVIEWED

Case 19. Ships and sanctions: watching trade from space

Task	track cargo and spot evasion at sea
Type	supervised and self-supervised
Model	vision models on satellite and ship data
Data	ship positions, radar images, ownership
Result	oil-flow estimates weeks ahead of trade data

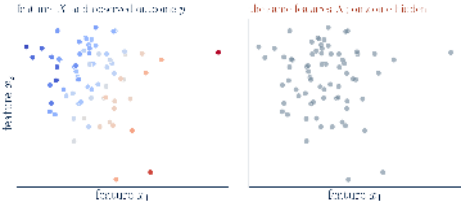
This became macroeconomic news after 2022, when sanctions created a market for finding ships that **switch their transponders off**.

A commercial data product became a macroeconomic instrument in about a year.

Source: Commercial providers; no independent audit of accuracy. COMPANY SAYS SO

Case 20. ESG ratings: scoring with no answer key

Task	score a company on E, S and G
Type	not disclosed; largely judgemental
Model	proprietary, not published
Data	disclosures; **no outcome to check against**
Result	six raters agree only 0.38 to 0.71



Two credit-rating agencies correlate at **0.99**. A borrower either defaulted or did not. **Nothing ever settles an ESG score.**

Before asking if a model is accurate, ask what it would be measured against.

Source: Berg, Kölbel and Rigobon, *Review of Finance*, 2022. PEER-REVIEWED

The same patterns in Chinese finance

Pattern from the twenty cases	Where you will meet it
Real-time credit on thin files	consumer-credit engines inside the large payment platforms
Fraud scoring at network scale	payment platforms clearing very high daily volumes
Insurance pricing and claims	large insurers with in-house AI units
Recommendation and pricing	the major consumer-internet platforms
Text and document models	bank and broker research assistants

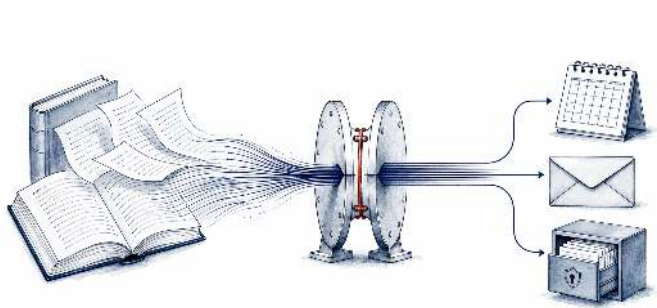
One honest warning. These firms publish very little that can be checked independently, so this slide names patterns, **not numbers**.

Apply the same five fields, and the same question: who measured it?

Part 3

Which model, and why

The answer is decided by the labels: how many you get, how fast, and how much each one costs.



Choosing a model family: the decision table

If your labels are	Use	Case from today
few, slow, contestable	logistic regression, GLM	credit scoring, insurance pricing
many, tabular, monthly	boosted trees	demand forecasting, claims
millions, arriving weekly	deep networks	card fraud
absent	clustering, networks	money laundering
the text itself	transformers	filings, contracts, calls
a sequence of actions	reinforcement learning	order execution

Nobody in these cases chose a model first. The label structure chose it for them.

Why fraud gets deep learning and lending does not

	Card fraud	**Consumer lending**
Labels per month	millions	thousands
Label arrives after	weeks	years
Cost of one error	small, reversible	large, and contested
Must you explain it?	no	yes, in the US and EU
Result	**deep network**	**logistic regression**

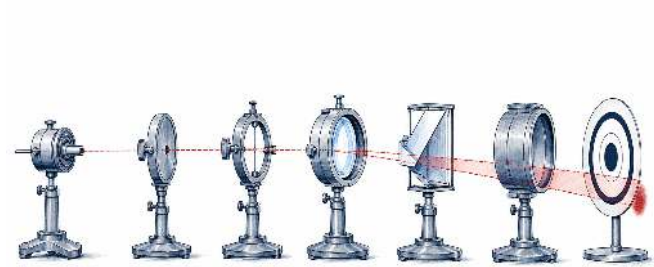
Same industry. Same firm, often. **Opposite technology**, for reasons that are economic and legal, not technical.

Count your labels before you choose your model.

Part 4

What fails, and why

Most failures are not statistical. They happen after the prediction, in the decision, the action, or the accountability.



Zillow: an ordinary forecast error, wired to a balance sheet

Task	estimate a house price, then buy the house
Type	supervised regression
Model	not fully disclosed
Data	listings, sales, tax records, photographs
Result	**$304m** written down; the business closed

Inventory write-down	∼$304m
Further losses guided	$240–265m
Workforce cut	∼25%

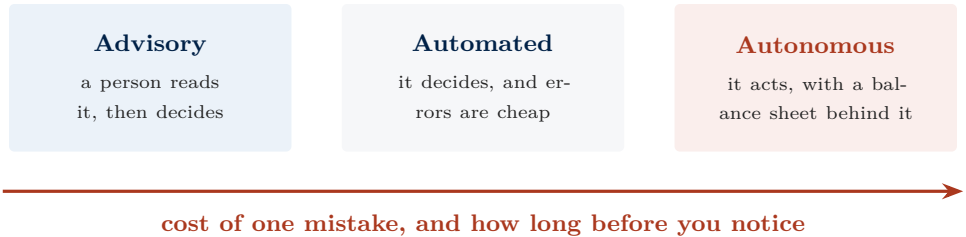
Zillow gave the cut as a percentage only. It published no job count; figures quoted elsewhere are estimates.

The pricing error was a few per cent, which is ordinary. The model was **wired straight to a purchase pipeline**.

The forecast error was normal. The wiring turned it into a company-ending loss.

Source: Zillow Group, Q3 2021 results and wind-down announcement, 2 November 2021. COMPANY SAYS SO

The question to ask of any AI system: how far can it act?



The **same model** can sit at any of the three. Nothing in the mathematics tells you which one you are looking at.

Card fraud is automated and safe. House buying was autonomous and was not.

Two famous failures with no model in them at all

	What happened	What failed
Trading firm, 2012	$440 million lost in 45 minutes	a software release procedure
Market-wide, 2010	prices collapsed and recovered in minutes	many correct algorithms interacting

The regulators' report on 2010 describes the trigger as **"a large fundamental trader"**. It **names nobody**. The name everyone repeats came from newspapers days later.

Learn to tell a wiring failure from a model failure. The fixes are completely different.

Source: SEC and CFTC joint report, 30 September 2010. REGULATOR OR COURT

The comparison that does not exist in your data

a third	a third	a third
improved the target metric	made no difference	made things worse

Ideas that passed Microsoft's internal review and were built. Only about a third improved the metric they were **designed to improve**.

A prediction tells you what will happen. It never tells you what your action changed.

Source: Kohavi et al., "Online Experimentation at Microsoft", Microsoft ExP team, 2009. COMPANY SAYS SO

We checked every number in this lecture. Half of them failed

36	51
confirmed against the original source	wrong, misdated, unverifiable, or never said

The split was not random. **Academic figures survived.** Vendor and press figures mostly did not. Examples: a fraud figure of "over 50%" that is really 32%. A fiscal-2023 number sold as "last year". A benchmark comparison **that was never run**.

This is the most useful habit this course can give you.

What fails, in one table

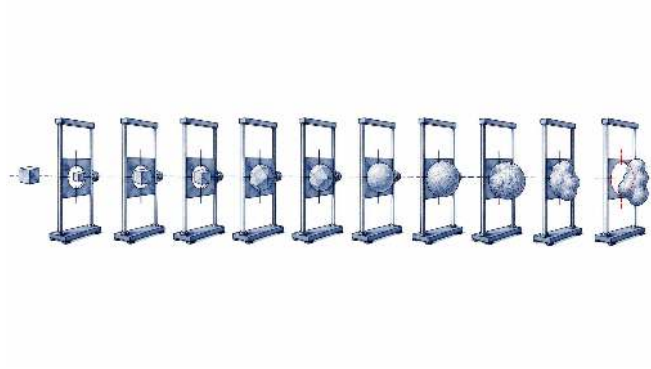
Failure	What goes wrong	Case
Wrong label	the model learns your past decisions	credit, hiring
Wrong wiring	a normal error becomes fatal	Zillow
No governance	nothing is watching at all	the $3.09bn fine
No comparison	you cannot show your action helped	telematics, Microsoft
Regime change	past relationships stop holding	nowcasting
No answer key	nothing ever settles the question	ESG ratings
Fluent but false	the text is confident and wrong	language models

Only one of these seven is a modelling problem.

Part 5

What is coming, and what is not

Some things are arriving quickly. Others will not be solved by making the models larger.



What is arriving in finance in the next few years

Direction	What changes	Blocked by
Agents, not chat	the model uses tools and takes steps	audit trails, permissions
Documents end to end	filings read, checked and summarised	verification cost
Private and on-site models	client data never leaves the bank	compute cost
Cheaper alternative data	satellites and sensors as standard	data licences
Regulated model governance	AI oversight becomes a named duty	shortage of qualified people

Notice the third column. **None of the blockers are modelling problems.**

The frontier in finance is governance and verification, not accuracy.

What making the models bigger will not solve

Problem	Why scale does not fix it
Causality	the comparison world is not in any dataset
Regime change	the future is not drawn from the past
No answer key	nothing ever reveals the right answer
Accountability	somebody must sign, and a court will ask who
Data you cannot get	the constraint is legal, not technical

Every one of these appeared in Part 2, in a different industry, wearing a different name.

These five are where a finance graduate is worth more than a larger model.

How much will AI add to the economy? Two serious answers

	Acemoglu	Goldman Sachs
Ten-year effect	under 0.53% added productivity	7% of global GDP
Method	builds up task by task	assumes broad substitution
Reviewed by	journal referees	the bank itself

Same technology. Same decade. **An order of magnitude apart.** Neither author is foolish; the methods differ.

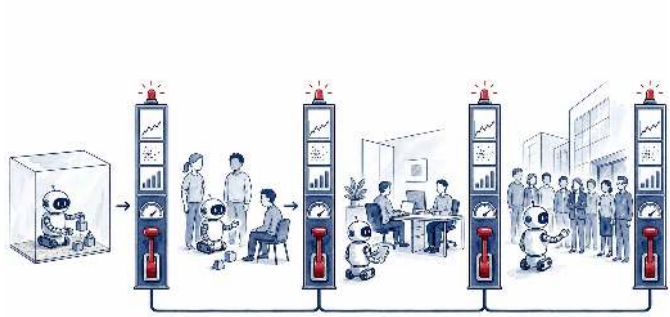
Nobody can sell you certainty about the aggregate, at any price.

Source: Acemoglu, NBER 32487 and *Economic Policy*, 2025 PEER-REVIEWED; Briggs and Kodnani, Goldman Sachs, 2023 PROJECTION.

Part 6

How to invest your time

Ordered by what you cannot undo. The first twelve months cost approximately nothing.



The graduate market, measured

Using payroll records from the largest United States payroll provider:

- **16% relative decline** in employment;
- for workers aged **22 to 25**;
- in the most AI-exposed occupations.

Not a survey. Not a forecast. Payroll data, about people your age.

-16%

relative employment, ages 22–25, most exposed occupations

The concern is real and measured. What follows is what to do about it.

Source: Brynjolfsson, Chandar and Chen, Stanford Digital Economy Lab, November 2025. PEER-REVIEWED

But notice where these tools actually help

	Where AI augments	Where AI automates
Who gains	beginners most (Case 13)	nobody in the role
Effect on you	you are productive sooner	the entry job disappears
Example	support, research, drafting	simple data extraction

Hiring data shows the same split: software postings recovered, but **71% of the increase was senior roles**.

Aim for work where the technology needs a person, not work it replaces.

Source: Indeed Hiring Lab, July 2026. COMPANY SAYS SO

The order matters more than the list

Months	What	Why here	Cost
0–3	SQL, and the command line	no substitute; nothing else works without them	free
3–9	Statistical learning, on real data	the standard free textbook, with its labs	free
6–12	Causal inference	your advantage over a computer scientist	free
9–18	One AI coding tool, one rule	only after you can do the work without it	$10–20/mo
12–24	One credential, for a named job	only after cheap practice has tested the direction	four figures

Running total through month twelve: approximately zero.

Why that order, and not the reverse

The common mistake	Why it happens
Buying the expensive course first	paying feels like acting, and a receipt is visible
Choosing it before you know the field	the irreversible purchase is made when you know least
Collecting certificates	they are easier to finish than a project

You already know the right principle from finance: **exhaust the cheap reversible options before the expensive irreversible one**.

If you will not finish a free course, a paid one will not save you.

The mathematics you need, and what you already have

Topic	Why	Status
Matrix algebra	every model is matrix multiplication	you have it
Probability, likelihood	how models are fitted	you have it
Partial derivatives, chain rule	how networks learn	one week
Eigenvectors	PCA and factor models	one week
Out-of-sample testing	the only test that matters	**new habit**

The gap is not knowledge. It is a habit. Econometrics trains you to judge a model **in** sample; machine learning only counts **out** of sample.

You are two weeks of mathematics away, and one habit.

Coding: SQL first, Python second

Skill	Used for	Honest note
SQL	getting the data at all	most requested, least taught
Python	pandas, scikit-learn	you need less than you fear
Git	working with others	non-negotiable in any team
Command line	running anything real	one afternoon
Excel	still everywhere in finance	do not be snobbish

Nobody is hired to build models on day one. **They are hired to get the data into a usable state**, which is mostly SQL.

If you cannot join two tables, no certificate will rescue you.

What to pay for, and what not to

Option	Cost	Verdict
Free textbooks and labs	$0	start here, always
Video course subscription	~$49/mo	fine if you finish it; most do not
Interactive platforms	~$12–25/mo	good drills, does not transfer to a real project
AI coding tool	$10–20/mo	yes, after month nine
Risk certification	~$2,000	only if the target job asks for it
Quant finance certificate	five figures	not before a paid role tests the direction
Bootcamp	four figures	market has contracted; check outcome reporting

Buy a credential for a job you have already named. Never to feel safer.

Source: Provider pricing checked August 2026; treat all prices as current then, not now. COMPANY SAYS SO

Using AI tools: what the evidence actually shows

+24%	−19%	+20%
what they expected	what was measured	what they still believed after

Sixteen experienced developers, on real work they knew well. They were **slower**, and could not feel it.

Write the first attempt yourself. Not as a virtue, but to keep a measurement.

Source: METR, July 2025 PEER-REVIEWED. Sixteen people; interval +2% to +39%.

Use agents, not a chat box

	Chat box	Agentic tool
Where it works	one answer at a time	inside your actual project
What it sees	what you paste	your files, your data, your errors
What you learn	an answer	how the work is put together
Risk	wrong and confident	wrong, confident, and already applied

The step that matters is calling the model **from your own code**, not from a website. Then you can test it, log it, and evaluate it.

The skill is not prompting. It is checking, and knowing what good looks like.

Your first ninety days, and what to stop doing

Start

- SQL, until joins are automatic
- One real dataset, end to end
- One question you actually care about
- Put it where someone can read it
- Read fifteen live job adverts

Stop

- Collecting unfinished courses
- Buying certainty when anxious
- Learning frameworks before statistics
- Quoting numbers you have not checked
- Letting the model write your first draft

One finished project beats five certificates, and everyone in hiring knows it.

The four questions, one last time

Who measured it?	Against what?	Which errors matter?	Who profits if you believe it?

Every number in this lecture has a date on it and will go out of date. These four will not.

Ask them of every vendor, every employer, and every plan sold to you for your own life.

Including the plan from the person who has been talking to you for four days.

Sources

Credit	Durand, *Risk Elements in Consumer Instalment Financing*, NBER, 1941. Berg, Burg, Gombović and Puri, *Review of Financial Studies* 33(7), 2020, 2845–2897. CFPB Upstart disclosures, 2019.
Fraud and crime	Stripe Radar documentation. US Department of Justice, FinCEN, OCC and Federal Reserve, October 2024. Hong Kong Police, 2024.
Markets	Gu, Kelly and Xiu, *RFS* 33(5), 2020, 2223–2273. SEC and CFTC, *Market Events of May 6, 2010*, 30 September 2010.
Insurance	Spedicato, Dutang and Petrini, *Variance* 12(1). University of Pennsylvania telematics trial, *Accident Analysis and Prevention*, 2025.
Business	M5 accuracy competition, *International Journal of Forecasting* 38(4), 2022, 1346–1364. Kohavi et al., "Online Experimentation at Microsoft", 2009. ACM FAccT, 2025.
Text and language	Caldara and Iacoviello, *AER* 112(4), 2022, 1194–1225. Loughran and McDonald, *Journal of Finance* 66(1), 2011, 35–65. Hansen, McMahon and Prat, *QJE* 133(2), 2018, 801–870. Hendrycks, Burns, Chen and Ball, CUAD, NeurIPS Datasets and Benchmarks, 2021. EMNLP Industry Track, 2023.
Economics	Federal Reserve Bank of New York Staff Nowcast. Berg, Kölbel and Rigobon, *Review of Finance* 26(6), 2022.
Work and careers	Brynjolfsson, Li and Raymond, *QJE* 140(2), 2025. Brynjolfsson, Chandar and Chen, Stanford Digital Economy Lab, November 2025. METR, July 2025. Acemoglu, NBER 32487 and *Economic Policy* 40(121), 2025. Goldman Sachs, April 2023. Indeed Hiring Lab, July 2026.

Every figure was independently re-checked before this deck was written. Where a widely repeated number failed that check, the corrected version is the one taught.