Panel Discussion: AI and Finance

Peter Forsyth
(University of Waterloo)

Sjors Altemuhl
(De Nederlandsche Bank)

Fang Fang
(FF Quant Advisory)

Diederik Fokkema
(ADC Consulting)

Roger Lord
(Cardano)

Format

Three rounds of discussion

- Each participant has three minutes in each round
- The audience is invited to join in after each round

Topics

- Round 1:
  - ↪ A good/successful story about machine learning/AI in finance
- Round 2
  - ↪ A bad story about machine learning/AI in finance
- Round 3
  - ↪ Where are we going? Where will we see a big impact?

What worked for me?

> "Everything in finance is an optimal stochastic control problem." *Robert Ferstenberg, MD, Morgan Stanley*

Optimal control

- Low dimensions $\rightarrow$ HJB PDE
- High dimensions $\rightarrow$ PDE not feasible

Solution:

- Approximate the control as a function of state using a Neural Network
  - Expectation $\Rightarrow$ sampling
  - Use standard tools (Pytorch, SGD, etc.)

Neural Network $\Rightarrow$ high dimensional function approximator

$\hookrightarrow$ Validated against HJB solve in low dimensions

$\hookrightarrow$ Stable even with large number of parameters.

SjorsRound 1

‘Successful’ story – insurance pricing

- Dutch insurers increasingly use machine learning (ML) and AI for applications like:
  - **Automation and customer interaction**
    Increasingly using generic LLMs (ChatGPT, Claude, Copilot, etc.)
  - **Fraud detection, reserving and pricing**
    Already for some time, using ‘traditional’ ML models (random forest, gradient boosting, etc). Experiments using generic LLMs.
- For more details, see DNB’s recent publications:
  - *Verzekeraars en AI: DNB deelt nieuwste inzichten | De Nederlandsche Bank (21 Jan 2026)*
  - *AI bij verzekeraars: kansen en risico’s | De Nederlandsche Bank (27 March 2025)*

Insurance pricing

- Insurers traditionally (and still) use *Generalised Linear Models* (GLMs) to determine insurance premiums.
- GLMs however can’t match the flexibility of ML and AI models, to accurately model risks on a granular level.
- ‘Succesful’, because the flexibility of ML and AI models has important downsides:
  - Insurers want (and are required) to be able to explain how their premiums are determined
  - Premium differentiation not allowed based on certain factors (sex, race, etc)
  - Granular pricing could make some subgroups uninsurable

DeNederlandscheBank
EUROSYSTEEM

FangRou

PANEL DISCUSSION · IN THREE ROUNDS

AI in Finance

The Good, the Bad, and What's Next

Round 01

What Worked

Wins from AI in finance

Round 02

What Didn't

Where AI fell short

Round 03

Where We're Going

Blue-sky / speculative

A practitioner's view

01

ROUND ONE

What Worked

A productivity revolution: work that used to take a small team weeks now ships in days.

FangRound

ROUND 1 · WHAT WORKED

AI on the Job: A Productivity Leap

Where AI is genuinely earning its seat on a quant / finance team today.

Coding & Tooling

Boilerplate, refactors, unit tests, and data plumbing that used to absorb a junior week now take an afternoon.

Designing AI Architecture

Sketching pipelines, comparing model choices, and stress-testing trade-offs becomes a fast back-and-forth instead of a multi-day study.

Documenting & Communicating

Research notes, model cards, internal memos, and stakeholder summaries — drafted in minutes, polished in one pass.

Deriving Proofs & Specs

Algebraic checks, edge-case enumeration, and formula derivations — a useful, very fast second pair of eyes.

Bottom line: *one senior + two AI agents now ships in a week what used to take a small team weeks.*

Fang Roun

ROUND 1 · WHAT WORKED

The Productivity Math

Same deliverable, very different team shape. The compression is real — and it's already showing up on roadmaps.

BEFORE	AFTER
Small team	One senior + AI agents
4–5	**1 + 2**
people	human + agents
over several weeks	***within 1~2 weeks***

WHAT ACTUALLY CHANGED

The senior's job, redefined.

Less time: writing first drafts of code, docs, and derivations.

More time: framing the problem, choosing what's worth building, and reviewing what the agents produced.

Net effect: throughput up, headcount flat, judgment matters more than ever.

Illustrative team-sizing from a working quant/AI practitioner — not a benchmark.

DiederikRo

ADC AI transformation

MATHEMATICS AND MACHINES · LEIDEN UNIVERSITY · 28 MAY 2026 · 13:15–14:15

AI and Finance

Panel Discussion · Three rounds · Three stories

Round 1 · Success

Round 2 · Failure

Round 3 · Blue Sky

DiederikRo

ROUND 1 OF 3 · 3 MINUTES · THE GOOD NEWS

Leiden · 28 May 2026 | AI & Finance

AI in Finance: One Thing That Actually Worked

The story

Sceptical that LLMs could do rigorous, auditable model validation — CRR3, EBA guidelines, MRM committee grade.

I compiled the regulatory framework into **107 verifiable claims** across **10 gates**: data governance → Merton/KMV → t-copula → 1M-path Monte Carlo → Euler capital allocation. An LLM orchestrates the full pipeline and signs the audit report.

Why it worked

Unambiguous targets | Formal thresholds | Verifiable outputs

EU regulation has hidden mathematical structure. When you compile that structure into a type system, LLM errors get caught at gate boundaries — not in a board paper.

107	10	< 5 min
PROOF OBLIGATIONS	VALIDATION GATES	FULL PIPELINE RUNTIME

The gate sequence

G0 Data quality → G1 Merton/KMV PD
G2 PCA factor model → G3 t-copula (ν)
G4 Loss distribution MC → G5 P1 floor
G6 Granularity → G7 Euler allocation
G8 Stress testing → G9 Governance sign-off

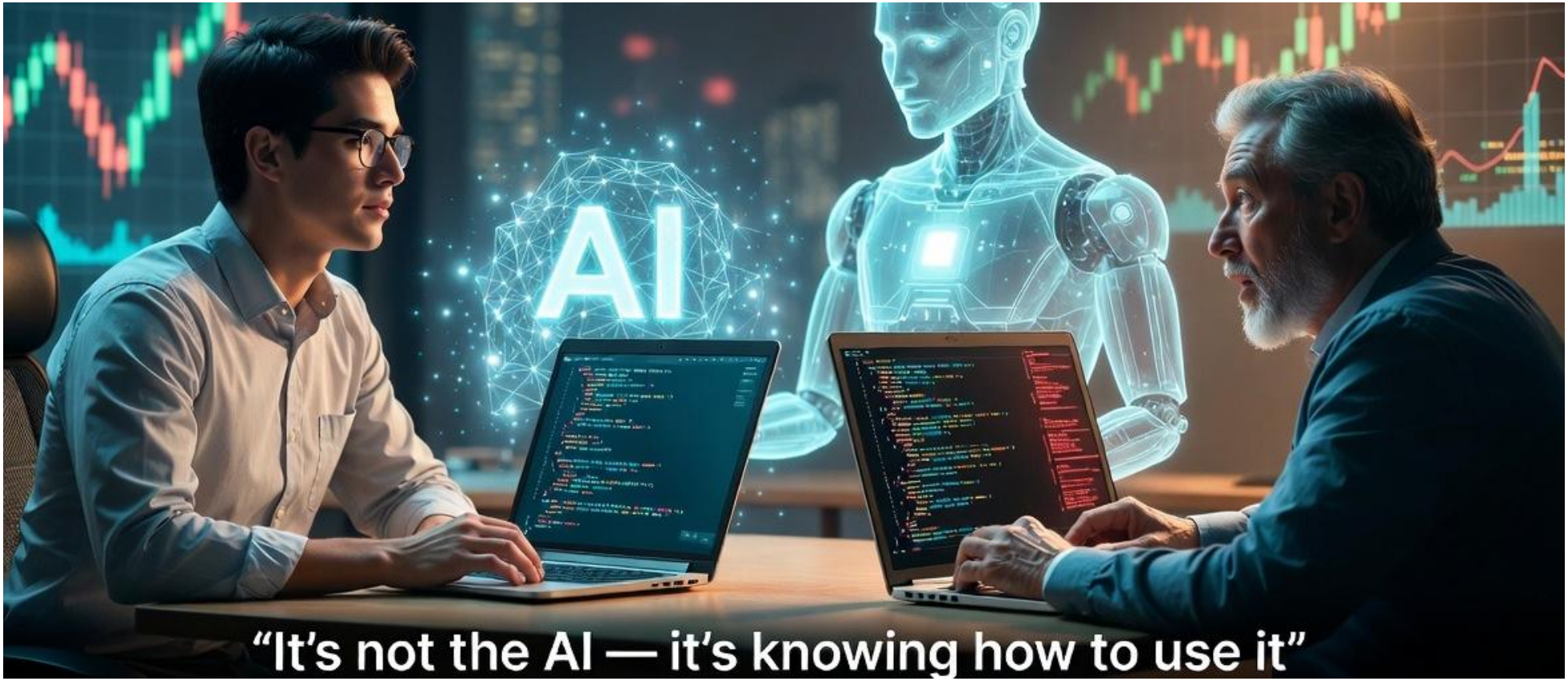
Punchline

When regulation is formal, AI becomes fast *and* safe. The LLM doesn't need to understand finance — it needs unambiguous targets.

RogerRo

AI as a multiplier, not a replacement

The interview assignment



Audience

What worked?

What is your good story?

What did not work

AI: seems to be able to do 3rd year undergrad math

- asked an AI a sophisticated graduate student question

AI claim $\Rightarrow$ known result

- Proof looked plausible
- 45 min reading proof, something did not look right

$\hookrightarrow$ Asked for references?

- Some references to me? (I did not do this)
- Other references were made up
- Another 1/2 hour, found error

Bottom line

- An LLM is trained to produce reasonable looking results
  - $\Rightarrow$ Does not mean its correct
  - $\Rightarrow$ Non-trivial to find errors

SjorsRound

Unsuccessful story – risk modelling

- **Personal opinion:** ML and AI most succesful in the following setting:
  - The application involves a large number of repetitive 'interactions' (questions to answer, movie recommendations to make, claims to assess, premiums to determine, etc.)
  - Individual results are relevant, not (only) aggregated results
  - Full understanding of results not paramount
  - Limited number of errors acceptable

Risk modelling

- Insurers use statistical, actuarial and financial models to quantify and manage their risks.
- For large insurers, capital buffers are determined directly based on these 'internal models'.
- Some experiments with machine learning and AI in risk models, but almost no actual use.
- My explanation:
  - Aggregated result most relevant (capital buffers based on total risk levels, not on risks for individual assets or insurance policies).
  - Full understanding paramount, and no margin for error. Supervisor (DNB) needs to approve models.

02

ROUND TWO

What Didn't Work

AI is a brilliant librarian — and a confident, plausible-sounding inventor. The gap is where harm hides.

FangRound

ROUND 2 · WHAT DIDN'T WORK

Querying ≠ Inventing

The same model that brilliantly recalls existing finance literature confidently fabricates novel results.

STRONG · Recall & Synthesis

- **Summarize a 30-page paper**
- **Pull the standard derivation of a known result**
- **Translate code between languages**
- **Reproduce a textbook proof**

Why it works

The answer already exists somewhere in the training data. The model is interpolating across known territory — and a human can verify it against a source.

WEAK · Genuine Invention

- **Derive a genuinely new pricing model**
- **Invent a novel signal nobody has tried**
- **Prove an unproven mathematical claim**
- **Judge if a new idea actually works**

How it fails

Output is fluent, confident, and plausible — and quietly wrong. There is no source to check against, because the result is supposed to be new.

The trap: *we trust the new answer as much as the old one — because both come back equally polished.*

Fang Round

ROUND 2 · WHAT DIDN'T WORK

The Hidden Cost: Who Trains the Next Generation?

AI is cheaper than coaching juniors — and the skills it replaces in year one are the ones they need to become good seniors.

THE QUIET RISK

"If the AI writes the first draft,
who learns to write the first draft?"

Critical, independent thinking is built by struggling through bad first drafts. Skip the struggle and you skip the formation.

Where it bites first

The hiring math has changed

An AI agent costs less than the salary, training, and patience a junior requires. For finance teams under cost pressure, the on-ramp into the profession is quietly closing.

No juniors today, no seniors tomorrow

Seniors are made — slowly, expensively, by doing junior work. Skip the apprenticeship and in 10 years there are no seniors left to do the judging that AI still can't.

Teaching & mentorship

How do you teach derivation when a student can summon a polished derivation in seconds? Universities and trading desks both need new rules.

Every junior role replaced by an agent saves money today — and removes a future senior from the pipeline.

DiederikRo

ROUND 2 OF 3 · 3 MINUTES · THE BAD NEWS

⚠ Finding | AI & Finance

Confident, Fluent, Wrong

The story

One prompt, no scaffolding — model document plus CRR3 regulation. Back came 400 words: article citations, threshold values, a structured verdict. Flawless language. Airtight logic.

The problem

Two of the six cited paragraphs said something *subtly different* from what the LLM claimed:

- Article 180(1)(a): threshold stated as ≥ 20 observations; LLM said ≥ 5
- EBA GL/2022/06 §43: condition inverted — LLM got the direction of inequality wrong

Neither error was obvious. A junior analyst would have signed it off. It took 90 minutes per error to locate.

Why this is dangerous

`Hard to detect without the source text`

`Compounding — wrong premise → wrong model → wrong capital`

`Asymmetric — lazy reviewer accepts; careful one wastes hours`

The asymmetry

	LLM CORRECT	LLM WRONG
Reviewer trusts	✓ Fine	✕ Silent failure
Reviewer checks	Wasted effort	Found, but costly

Expected cost rises sharply once errors become hard to detect.

The mathematical analogy

A plausible-looking but wrong proof is worse than no proof. You cannot check an LLM's regulatory reasoning by inspection — you need to verify against the ground truth independently.

Punchline

A hallucination in a board paper is worse than silence. Fluency is not a proxy for correctness — and in finance the cost of that confusion is systemic.

RogerRound

Where the hype currently hits the wall

It works on my laptop vs. it works at the firm

Audience

What did not work?

What is your bad story?

Where are we going?

Many years of being a member of UofWaterloo pension investment committee

- Every quarter, receive a massive slide deck
- Graphs, colour pie charts, complex metrics
  - → Obfuscating the fact that active managers underperform the index

My colleague fed the recent 90 page slide deck into Chatgpt:

- Instruction: Please summarize in a few sentences

45 minutes later:

> "the portfolio's complexity has not added value over a passive implementation..."
> "The hidden risk behind the whole strategy is the asset-liability mismatch.."

LLMs can be a good bullshit detector!

SjorsRound

Future of AI in insurance

- Gradually, I expect more and more use of AI in insurance:
  - First the repetitive tasks
  - Not necessarily replacing humans, but producing results for humans to work with
  - 'Principle-based' regulation difficult for AI

How do you remain relevant?

- **Learn to work with AI**
  young people have an advantage!
- **Critical thinking and deep understanding**
  AI can present wrong results convincingly. Assessing the appropriateness of AI results, given the context, is where you must make the difference.

DeNederlandscheBank
EUROSYSTEEM

03

ROUND THREE

Where We're Going

If we get the guardrails right, the upside is enormous. If we don't, Round 2's risks scale with it.

FangRound

ROUND 3 · WHERE WE'RE GOING

If We Solve Round 2: The “One-Person Team” Era

Picture Round 2's risks managed — verified outputs, juniors who still learn judgment. The upside: one skilled operator does what a small team used to.

Diagram: one senior orchestrates a team of specialised AI agents.

WHAT THIS UNLOCKS

Faster iteration

Hours, not weeks, between question and tested answer.

Smaller, smarter teams

Headcount budget shifts toward judgment, not throughput.

Lower barrier to entry

Boutique shops can now compete on ideas, not infrastructure.

Premium on taste

When everyone can build, knowing what to build wins.

FangRound

ROUND 3 · WHERE WE'RE GOING

Closing the Loop on Round 2

Three pillars to keep AI's polished-but-wrong outputs — and the atrophy of judgment — from scaling with the productivity wins.

PILLAR 01

AI Detection

Real-time tools that flag synthetic content before it hits a customer, a market, or a research desk.

- Deepfake & voice-clone detection at the front line of KYC and customer service.
- Provenance & watermarking for research, trade memos, and approval documents.
- Anomaly models that catch "too-perfect" activity — the tell of a coordinated AI campaign.

Answers Round 2's "polished but wrong" problem at scale.

PILLAR 02

Regulation

Rules of the road. Without them, the firms that cut corners on AI safety win in the short run.

- Disclosure: when AI materially shaped a recommendation or decision, say so.
- Audit trails for AI-assisted approvals — who, what model, what version, what prompt.
- Liability that travels with the output, not just the human who shipped it.

Levels the field so honesty isn't a competitive disadvantage.

PILLAR 03

Human Pipeline

Even when AI is cheaper, firms still need to hire and grow juniors — or in 10 years no seniors are left.

- Treat junior development as long-run R&D, not a short-run cost line.
- Apprenticeship redesigned: juniors learn to challenge AI output, not just produce it.
- Curricula that teach when to trust AI, when not to, and how to tell the difference.

Keeps the junior-to-senior pipeline alive.

The bet: the firms — and societies — that build all three pillars in parallel are the ones that get the productivity upside ***without*** the harm.

DiederikR

ROUND 3 OF 3 · 3 MINUTES · BLUE SKY

Speculative AI & Finance

Regulation as Executable Code

The hidden structure of financial regulation

EU financial law is not just prose. It has a layered formal structure:

```
L0 Regulations   → what is law
L1 Policy        → governance principles
L2 Thresholds    → numeric bounds
L3 Claims        → proof obligations
L4 Specs         → executable gate DAG
L5 Evidence      → verifiable artefacts
```

A "compliant model" is a well-typed program. Non-compliance is a type error — caught at compile time, not in an MRM meeting.

The vision

LLMs become *code generators* discharging proof obligations against a formal spec. This is the **Curry–Howard correspondence** applied to finance: proofs ↔ programs, regulation ↔ types, compliance ↔ type-checking.

What changes

Regulators write **specs**, not prose
Banks run pipelines, not committee cycles
LLM errors caught at gate boundaries
Audit trail is machine-generated, hash-verified
Model risk becomes **software engineering**

The open problem

How do you compile ambiguous natural-language regulation into formal proof obligations without losing the legislator's intent? This is the research problem — and it is as much a math/PL problem as an AI problem.

Punchline

Not AI that *understands* finance — a type system for financial risk, with the LLM as proof assistant.

DiederikRou

GYREX proof obligation system

Diederik R. Fokkema, PhD

ADC Consulting

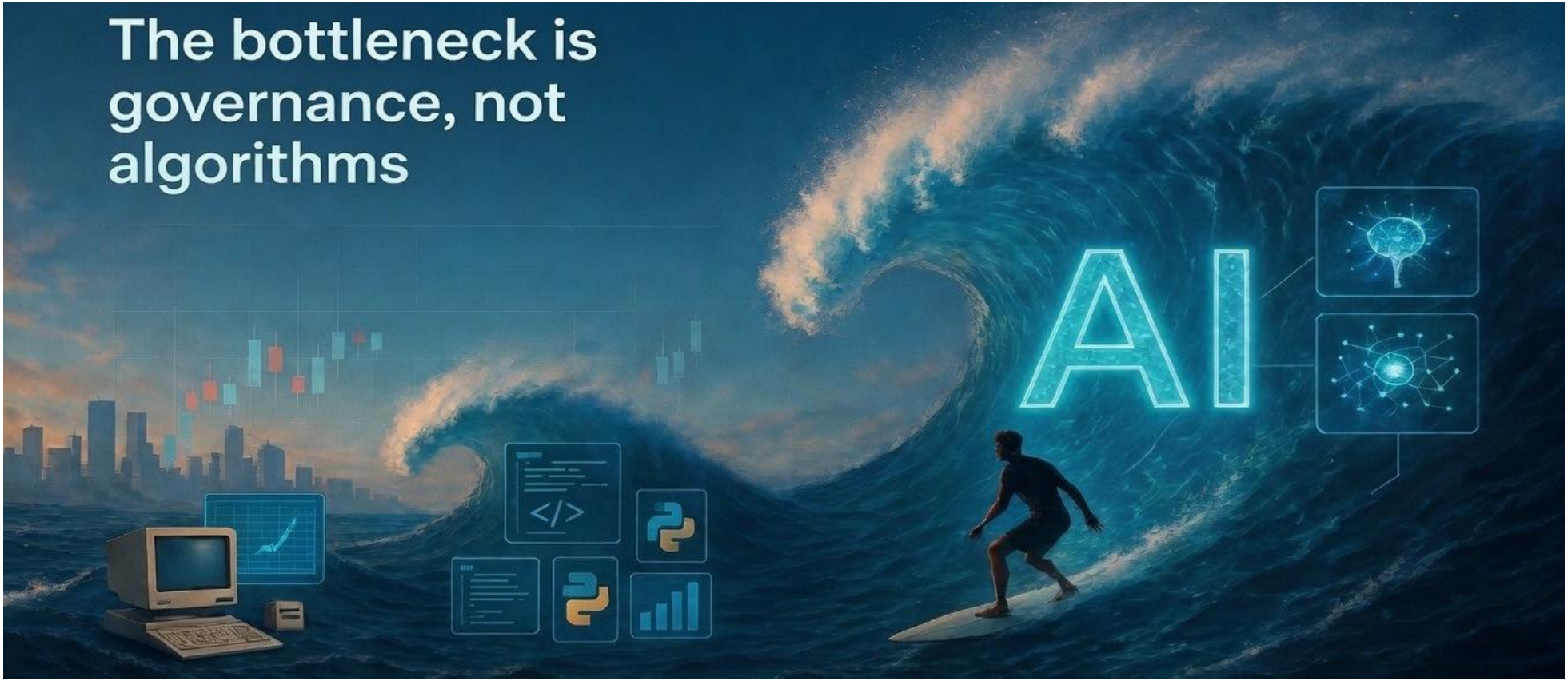
diederik.f@adc-consulting.com

+31 6 29 19 11 45

RogerRour

It’s just a scripting language

What Python adoption tells us about AI adoption



Audience

Where are we going?

Speculate: where will we see a big impact?