FUTURE OF FINANCE LAB

From AI-built models to confident decisions

AI + HUMAN EXPERTISE = BETTER DECISIONS

HIGHER CONFIDENCE
LOWER RISK
FASTER INSIGHTS
STRONGER BUSINESS

14. SPEED-TO-TRUST: CHECKING AI-BUILT MODELS LIVE AND POST-BUILD





with Lance Rubin & Maarten Bessems
Featuring Excel Analyzer

Thu 13 Nov | 10:00 AM AEDT | Free

PEOPLE + MODELS + A BRIGHTER FINANCIAL FUTURE

MODEL CITIZN

CONTENTS

Welcome to the Finance Lab and the AI-Powered Accountant

01

Quick introductions to the session outline

Who’s here, what we’ll cover and how the hour runs

02

Overview of today’s focus

Foundations, then Speed-to-Trust with Maarten Bessems: live and post-build checks

03

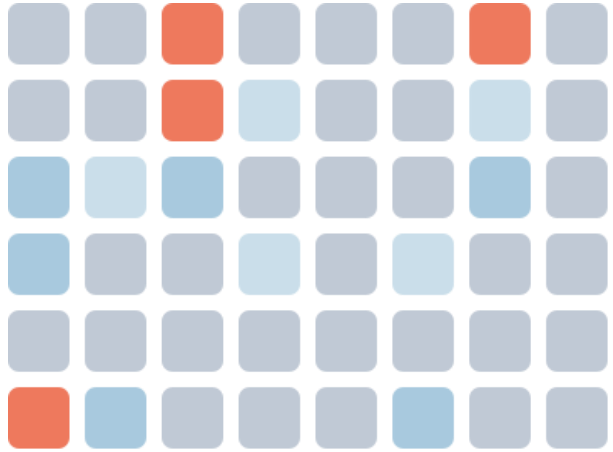
Discussion with our guest

Your questions on trusting, reviewing and fixing AI-built models

04

Wrap up and preview of the next session

Next steps, resources and what’s coming in the Lab



PART 1

Foundations and Concepts

The frameworks every Lab session builds on



HACK: the framework we apply

H



Hygiene

Clean, structured, standardised data is the prerequisite. Without hygiene, AI amplifies mess and produces confidently wrong output.

A



Automation

Once data is hygienic, automate the repetitive. Order matters: never automate before you clean.

C



Capability

The technical depth, technology assessment skills and governance awareness to operate the tools and verify the outputs.

TODAY: VERIFY THE OUTPUTS

K



Knowledge

Context, storytelling and willingness to unlearn: the human edge AI cannot replicate.

Sequence: H → A → C → K. Not parallel. Not optional. Not skippable. Hygiene and Automation are the environment; Capability and Knowledge are the practitioner.

Session #14 – Speed-to-Trust with Maarten Bessems

TODAY

1. Excel fluency

When all else fails it will be there to rescue you, but you need to know how.

2. Process & data

A process and data mindset, so you know how the pipes work before you automate them.

TODAY

3. Assess tech

The ability to assess technology, since not all apps are created equal.

TODAY

4. AI concepts

AI fundamentals: understanding hallucinations, tokens and guardrails.

5. Growth mindset

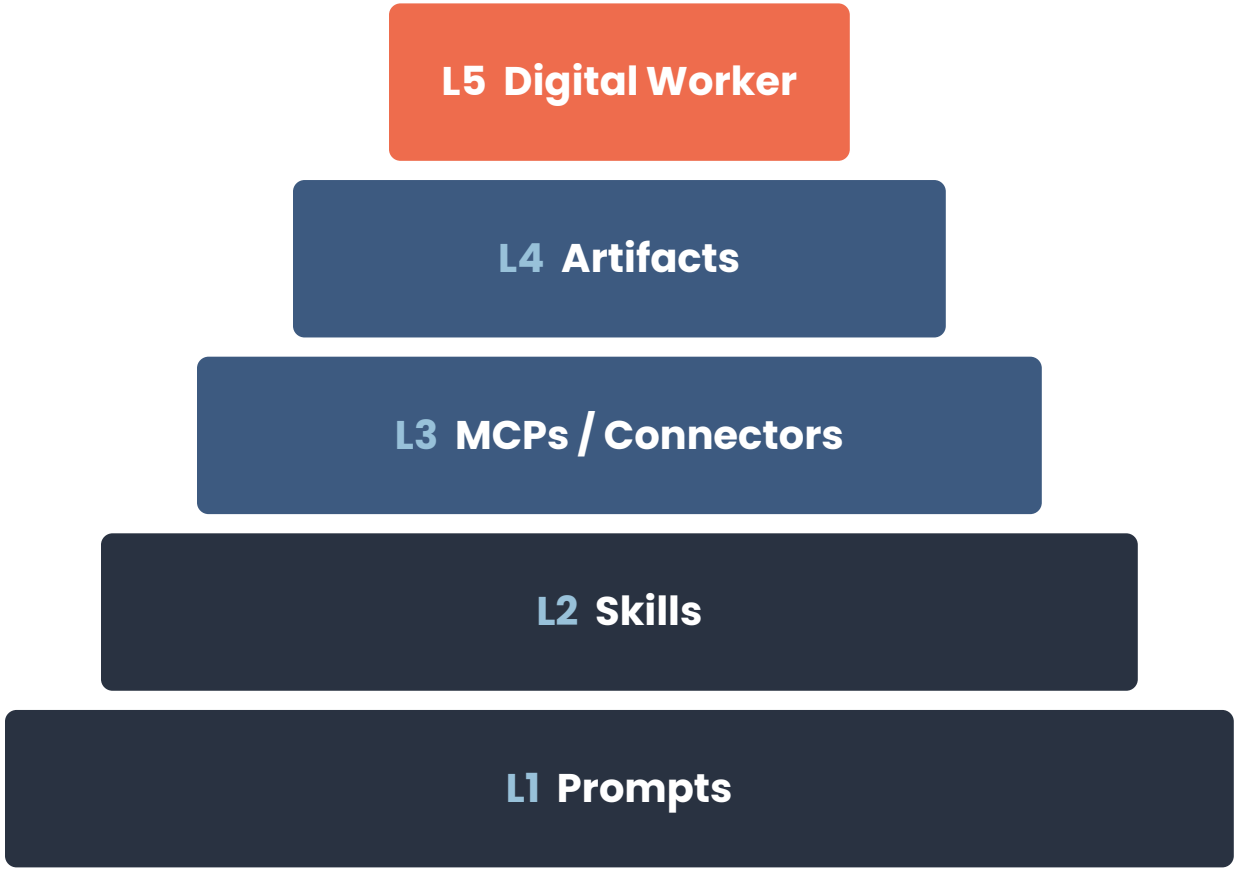
Where every failed workflow is a learning curve.

6. Human skills

Empathy, storytelling, influence.

Pillars 1–4: hard / technical · Pillars 5–6: soft / human · Today leans on 1, 3 and 4

The abstraction pyramid



The skill of skills

Skills + connectors made usable

Skills wired to real systems

Repeated prompts, crystallised

The atoms of every interaction

Each layer is built from the one below it. You are starting at the foundations.

2 critical concepts to remember

1 System thinking

- Understand interconnections, not isolated parts
- Think end-to-end process and what you want to achieve upfront, before you do anything
- Know what outcomes and outputs you expect before they're produced, and prompt accordingly
- Frameworks are key to system thinking
- Abstract from task level to orchestration level

2 Probabilistic vs deterministic

- Why AI behaves differently from Excel, and why that changes everything
- Accountants want accuracy; AI hallucinates. That's why you get AI to build formulas you can check, not numbers
- Today: ExcelAnalyzer is deterministic (same scan, same result). Claude is probabilistic. Speed-to-trust needs both

Together, system thinking and probabilistic vs deterministic are key. AI will make mistakes, so we have to put the systems in place to reduce that risk.

What it looks like under the hood

Daily Briefing

The morning brief that pulls email, calendar and client focus into one view

Skills Library

280+ skills encoding how we model, review, report and teach

AI Projects Hub

The command centre: every project, its context and its outputs

AI Evals

How we test each new model release before we trust it with client work

Today's benchmark is one of those skills: an L2 skill running on L3 connectors (Claude + Excel), fed by ExcelAnalyzer evidence.

Maarten Bessems



CURRENT ROLE & FOCUS

Founder & CEO, Spreadsheetsoftware

The Maastricht team behind ExcelAnalyzer, ExcelCompare and Gridbox.

Focus: making spreadsheet models auditable, with live error overlays, risk scoring and review inside Excel, for human-built and AI-built models alike.

BACKGROUND

Maarten Bessems is the Founder and CEO of Spreadsheetsoftware, the team behind ExcelAnalyzer: an Excel add-in for finding errors, mapping formula logic and reviewing models. He has spent two decades building and auditing Excel models.

Capabilities

Ten years as a quantitative consultant and senior financial modeller at the Netherlands' largest real estate company. He has built, reviewed or audited more than 1,000 spreadsheet models across real estate, private equity and banking. MSc Financial Mathematics, Maastricht University; founded Spreadsheetsoftware in Maastricht in 2011.

Leadership philosophy

Review matters more in the AI era, not less. Let AI draft the numbers, but keep people in control, with evidence you can see cell by cell before anyone relies on the model.

linkedin.com/in/maartenbessems · spreadsheetsoftware.com

Discussion questions — Speed-to-Trust

1. Would you rely on an AI-built model without human review, and what would it take to trust one?

2. What are the most common errors you find when reviewing models, and has AI changed them?

3. Should the check happen live while AI builds, after the build, or both?

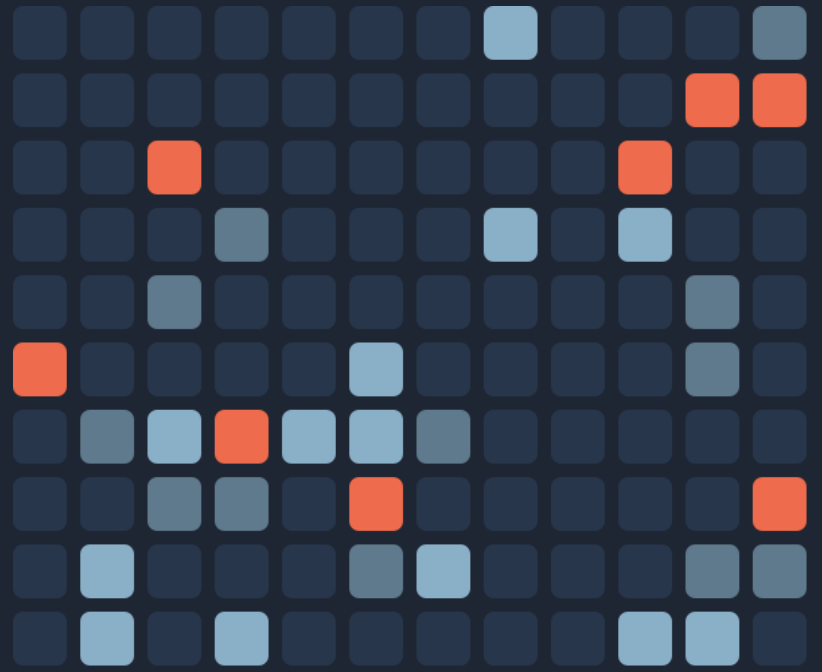
4. What can ExcelAnalyzer see that Claude or Copilot can't, and what can AI add that a tool can't?

5. Can you tell when a model was built with AI, and does it matter if it's right?

PART 2 · WITH MAARTEN BESSEMS

Speed-to-Trust

Checking AI-built models live and post-build

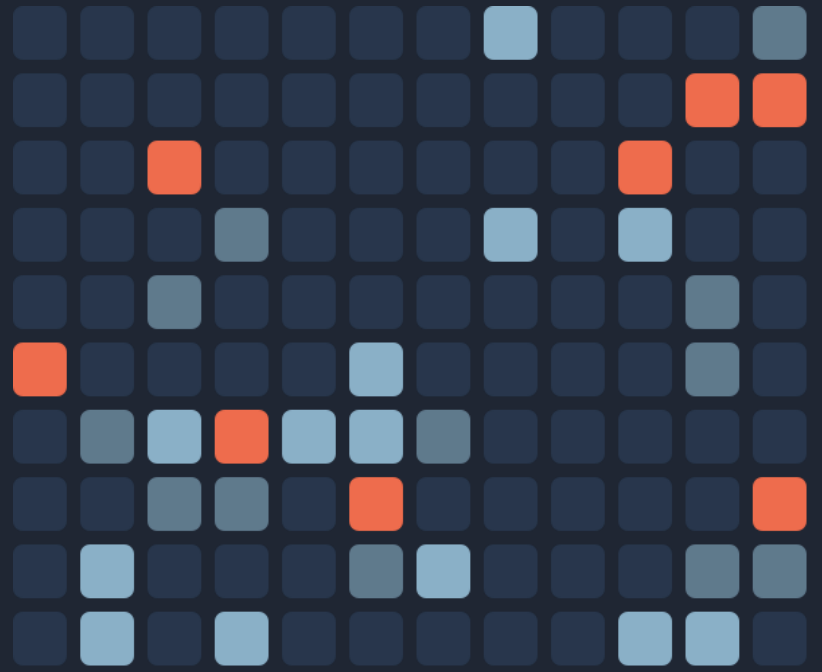


Disclosure: Model Citizn is an ExcelAnalyzer customer and uses the tool extensively.

PART 2 · LIVE CASE STUDY

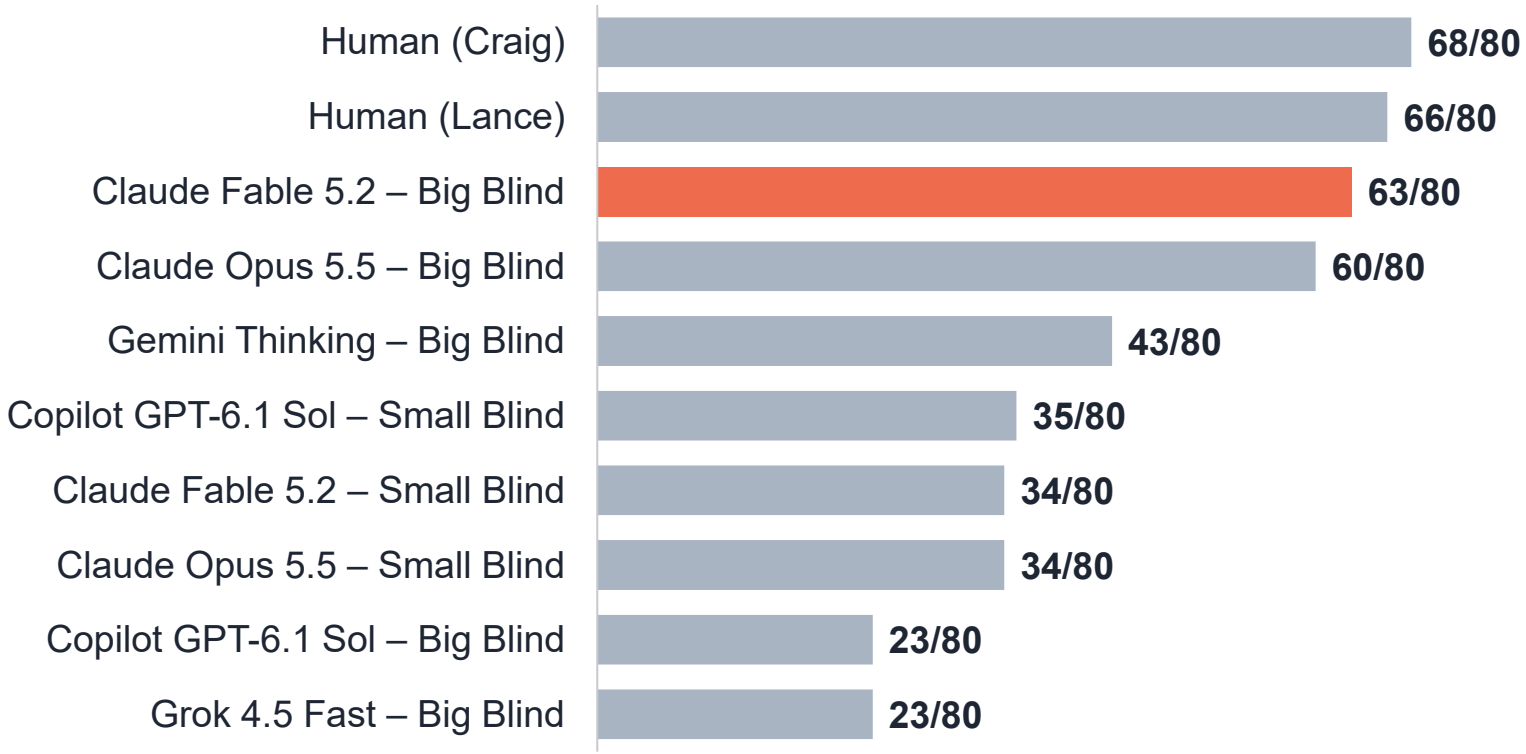
Case Study 5

8 AI models vs 2 humans, one brief



Skilled AI scored 60–63, humans 66–68, cold tools 23–43

Quality score after audit caps, by build (out of 80)



SAME BRIEF, THREE TIERS

- Human references: 66 and 68 out of 80
- Claude with EXL Cloud skills, Big Blind: 60 and 63
- Cold tools, Big Blind: Gemini 43, Grok 23, Copilot 23

WHAT MOVED THE SCORE

- The harness: skills and guardrails, not the model name
- Copilot Big Blind paused at its approval gate and built no statements

Big Blind: build from brief. Small Blind: rebuild from Formula Map. Claude: cloud workspace (skilled). Copilot: Excel. Gemini, Grok: web apps.

Source: Comparison model file v1.203 – Quality Scoring, Audit Results, Accuracy Check (47/47 validation checks PASS)

Operating 99.7% accurate, balance sheet only 6.0%

99.7%

Operating accuracy vs the human reference for Claude Fable, Claude Opus and Gemini (revenue and EBITDA, FY26–30)

6.0%

Balance sheet accuracy at Dec-30 for the same Claude builds. The two human models agree only about 70%

0 of 5

Big Blind AI builds that passed all three human-band tests: equity IRR, closing cash and net assets

Right engine, unfinished answer. Copilot GPT-6.1 Sol – Small Blind was 100% accurate on operations but 0% on valuation (66.7% overall): it stopped at an approval gate, leaving equity value at –$0.74m vs $39.8m and an IRR that did not calculate.

Source: Comparison model file v1.203 – Quality Scoring, Audit Results, Accuracy Check (47/47 validation checks PASS)

AI builds in minutes. Trust hasn't moved.

0 of 63

FMI Global Leaders Council members who would rely on an AI-built model without human review

14%–25%

Range of AI-built equity IRRs in our 12 Nov eval of 8 builds, against a human benchmark of ~21% (September: 13%–158%)

1 of 8

AI builds in that eval that passed all three accuracy tests against the human models: IRR, cash and net assets

98% correct is still 100% wrong. Clean isn't correct. The bottleneck is no longer the build; it's the review. Today: how to make that review fast enough to keep up with AI, live and after the build.

Eight AI builds, one brief: who matched the humans?

Model	Task · harness	Accuracy tests	Quality /80
Claude Fable 5.2	Small · skilled	0 of 3	34
Copilot GPT-6.1 Sol	Small · cold	IRR #NUM!	35
Claude Opus 5.5	Small · skilled	0 of 3	34
Claude Fable 5.2	Big · skilled	0 of 3	63
Claude Opus 5.5	Big · skilled	0 of 3	60
Gemini (Thinking)	Big · cold	0 of 3	43
Grok 4.5 Fast	Big · cold	0 of 3	23
Copilot GPT-6.1 Sol	Big · cold	0 of 3	23

THE HUMAN BAR

- Equity IRR 20.7% to 21.1%
- Closing cash ($1.88m) to ($2.14m)
- Net assets $3.58m to $3.61m
- Two EXL Cloud human models, scored 66 and 68 out of 80

SINCE SEPTEMBER

- IRR range narrowed from 13%–158% to 14%–25%
- First Gemini and Grok runs

Big Blind: build from the brief. Small Blind: rebuild a stripped model from its formula map. Claude ran in the Claude app's cloud workspace, Copilot in Excel, Gemini and Grok in their web apps. Skilled: Claude with EXL Cloud skills. Cold: no skills.

Check it live: six overlays, one click each

Formulas

Colours inconsistent, hardcoded and unusual formulas across rows and columns.

Dependents

Shows which cells feed others, so you can see what a change will touch.

Number formats

Flags mixed or inconsistent formats, including numbers stored as text.

Data validations

Marks input cells that have validation rules, and those missing them.

Locked cells

Shows protected cells, and formulas left open to accidental edits.

Conditionals

Reveals conditional formatting rules, including ones you didn't know were there.

ALSO IN THE BOX

- Realtime: flags broken formulas and inconsistent values as they appear
- Hidden rows, columns and very hidden sheets exposed
- Inter-sheet and external link mapping
- Panel counts and groups every finding
- Risk score for the whole model, plus a one-click audit report (PDF)

Overlays don't change your workbook. Run them while Claude or Copilot is building, and you see problems as they land, not three weeks later in review.

AI describes. ExcelAnalyzer locates.

DESCRIBES
Claude / Copilot

- Reads the model and explains issues in plain English
- Reasons about intent, logic and fit to the brief
- Probabilistic: can skip cells it never looked at, and answer differently each run

LOCATES
ExcelAnalyzer

- Pinpoints every issue cell by cell, with counts and denominators
- Deterministic: same scan, same result, every time
- Can’t tell you whether the logic answers the brief

TRUSTS
Together

- ExcelAnalyzer produces the evidence
- Claude scores, explains and plans the fixes
- A human reviews and signs off

Benchmark it: 175 criteria, scored on content

175

best-practice criteria, drawn from:

- EXL Cloud
- SSRB
- FAST
- Modano
- ICAEW
- FMI

Every run is tagged EA or non-EA, so you can see what ExcelAnalyzer evidence adds.

1. **Ask the EA question**
   Has ExcelAnalyzer been run on this model? The answer is recorded on the run.
2. **Collect read-only evidence**
   Inventory, checks, 3-way ties, hardcodes and consistency (with denominators), plus the docs and navigation content.
3. **Score every criterion 0–4**
   Evidence quotes sheets, cells and counts. “The sheet exists” is not evidence above a 2.
4. **Calibrate both ways**
   No inflated 4s, no placeholder scores, no penalising one issue under several criteria.
5. **Log the findings**
   Every Critical or High criterion scoring 2 or less goes to the findings log, ready to plan.

Spot the AI: signs a model was built with AI

STRONG

- Blank Author but a human lastAuthor in file properties
- Chat phrasing left in notes: “User confirmed”, “This is the fix”
- A templated note on every row, or notes citing another file’s cells
- One identical long formula stamped across a whole grid

MEDIUM

- Stamped guard skeletons around formulas
- Numbered method and “points to confirm” write-ups
- Uniform hedging in every note
- Overstated claims such as “every cell is a formula”

WEAK

- En dashes, arrows and superscripts
- Blank margin columns and gridlines off
- File properties that predate the content
- A careful human can produce all of these



It’s a likelihood, never proof, and AI-generated isn’t the same as wrong. Use it to decide where to check harder: in-app agent, code sandbox or pasted from chat each leave different traces.

Five checks before you trust an AI-built model

1. **Balancing isn't proof** A 100% tie can just mean it copied the answer.
2. **Hunt for hidden hardcodes** AI plugs numbers to force a result. Check the trail.
3. **Confidence is not correctness** It sounds certain even when the logic is wrong.
4. **Audit before you ship** Demand a formula trail on every single number.
5. **Keep a human on the logic** Own the structure and the judgement calls yourself.

Start · Own it

Middle · Supervise

End · Delegate

Set the structure and logic yourself. Let AI draft formulas, but audit every cell. Hand off reporting, commentary and board packs: that's where AI wins.

PART 3

Take Away – Next Steps

What to do with today, and where to go next



GUEST OFFER

Get 50% off ExcelAnalyzer

A special offer for attendees of this session

~~€800~~ **€400**

per license, per year
(max 2 licenses)

Until November 27, 2026

01

Reserve your spot now

Submit the form before November 27, 2026.

spreadsheetsoftware.com/speed-to-trust

02

Try it free for 10 days

The new version is ready on December 1, 2026.

03

Decide after the trial

Buy with your reserved 50% discount or don’t.

TAKE AWAY

Next steps



Share it

Share thoughts and experience publicly if you are comfortable, and invite colleagues and friends.



Catch the replay

Recording and contents will be shared in the Knowledge Hub, on Eloquens, on YouTube and on our Spotify podcast.



Get ready to practise

Future sessions share modelling examples, data and workbooks you can practise on and integrate AI into. Sessions remain FREE.



Coming to EXL Cloud

Skill 21, best-practice benchmarking, is in testing: run it with ExcelAnalyzer evidence or without.

TAKE AWAY

Your resources

1

Knowledge Hub
modelcitizn.com/knowledge-hub

2

Modelling Shop
modelcitizn.com/shop

3

Eloquens
eloquens.com/channel/lance-rubin

4

Maven
maven.com/modelcitizn

5

EXL Cloud
exlcloud.io

6

Future of Financial Modelling – YouTube
Watch on YouTube

What to expect and when

LIVE + ASYNC
Sessions 1–2
Modules 1–2

Foundations, formulas, PDF extraction and analysis

1 hr each + 5–7 hrs exercises

ASYNC
Session 3
Modules 3–6

Financial control, management reporting, source mapping, workflow analysis

5–8 hrs

ASYNC
Session 4
Modules 7–10

Structuring messy workbooks, validation and audit, data visualisation (Power BI + Excel), skills MCQ

4–7 hrs

LIVE
Session 5
Capstone

Build real workflows

7–9 hrs

LIVE
Session 6
Future of AI

Top 3 workflows and awards

1.5 hrs

~20–32 hours over 8 weeks · cohort-based

Now also self-paced: start any time, US$499, certification included, with Lance reviewing every project.

Lightning & Lab sessions: 1 hour, with guests and experts you can learn from.

Thank you

STAY CONNECTED

Lance Rubin
www.modelcitizn.com · linkedin.com/in/financial-modelling/

Maarten Bessems
spreadsheetsoftware.com · linkedin.com/in/maartenbessems

Recording on YouTube: The Future of Financial Modelling

