AFP FP&A SERIES

A Complimentary Virtual Event for Finance Professionals

Sponsored by ramp

The Human Layer of Owning AI Output

Anna Tiomina
Founder, Blend2Balance

AFP FP&A Series
August 26, 2026

ABOUT THE TRAINER

Anna Tiomina

- **Former CFO at Sandoz and Softeq.** 15+ years building finance operations at high-growth companies in the US and Europe

- **Founder of Blend2Balance.** I train finance and operations teams to use AI and advise finance leaders on getting it adopted across the team.

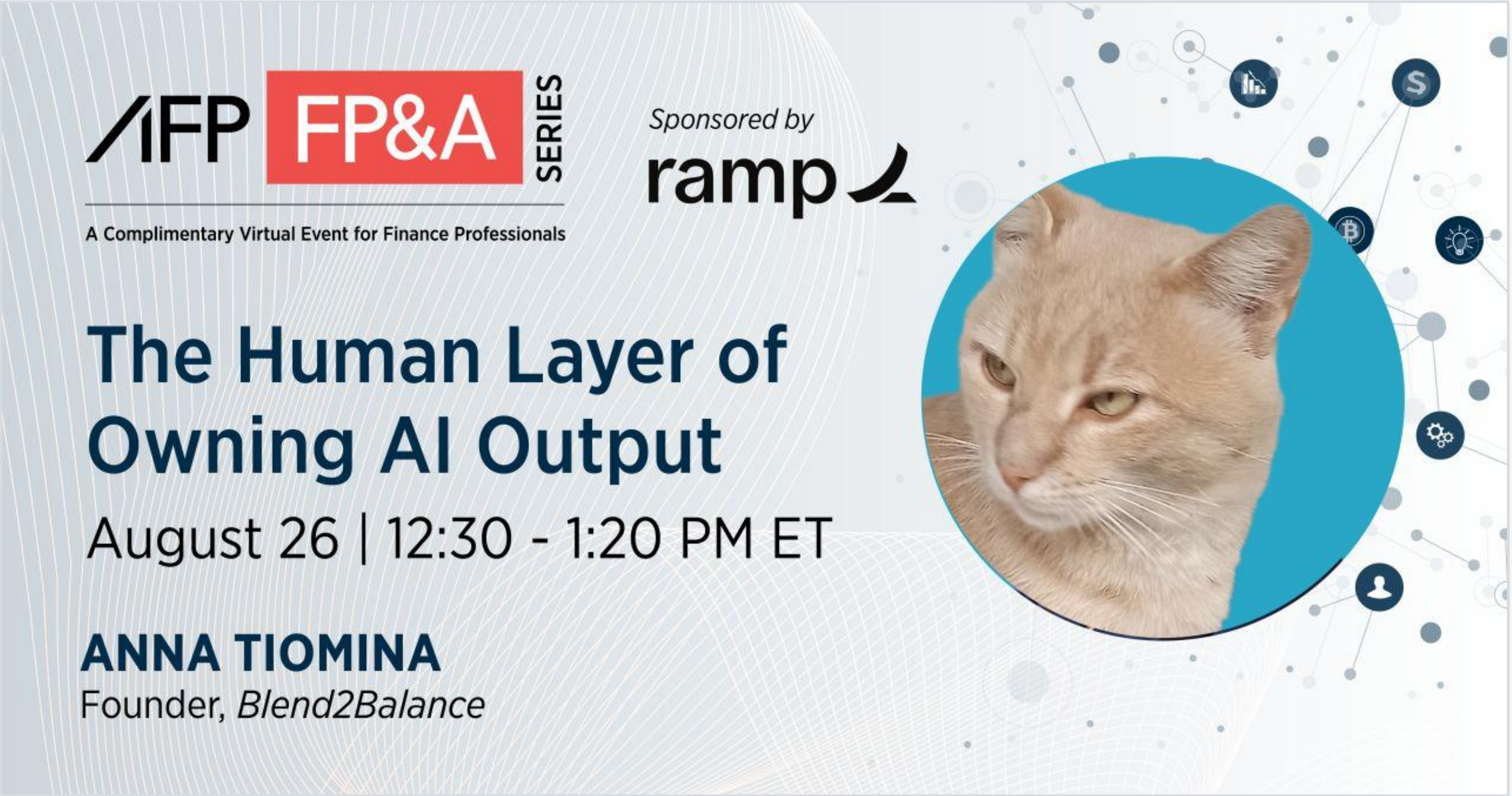
- **1,000+ finance professionals trained,** across PE-backed companies, startups, SaaS and professional services

Balanced AI Insights

Weekly, for finance leaders. Scan to subscribe

newsletter.blend2balance.com/subscribe

What is the funniest AI output you have ever received?



Mine is on the left.

I asked for a title slide with my headshot. Both files were in the same folder, and I never said which one to use.

Put yours in the chat.

The errors that get through look correct

An obvious error

The cat gets caught by everyone in the room, so it costs nothing.

An expensive error

The numbers tie, the driver sounds right, and nobody opened the source.

AI is already in our numbers, and we know it gets things wrong.
The question is where to trust it and how to verify it without losing more time than it saved.

What we will cover today

01 Three types of AI work in FP&A

02 The check that matches each type

03 Making the check affordable, and owning the output

04 Agents: where do the old controls stop working?

Three types of AI work in FP&A

01

Automation

AI runs the process.
You set the controls.

02

Accelerator

AI produces a draft.
You check it.

03

Thought partner

AI helps you think.
You decide.

The failure is different in each, so one review standard across all three does not hold.

Which type of AI work takes up most of your time today?

POLL 1

- **A** Automation. The process runs and I set the controls
- **B** Accelerator. It drafts and I check
- **C** Thought partner. It helps me think
- **D** None of these yet

What determines the kind of error you can expect

Deterministic step

One right answer for a given input.

Fails by breaking, or by drifting away from the business it was built for.

Probabilistic step

Generated judgment or language.

Fails by producing something plausible.

Most FP&A workflows contain both.

Misclassify a step, and the controls you build on it are wrong from the start.

Type one: Automation

Where it shows up

- Actuals load and mapping
- Budget versus actual roll-up
- Threshold-based variance flagging
- Source-to-model reconciliation
- Scheduled pack refresh

What goes wrong

- Rules stop matching the business
- The output still looks clean
- Drift you cannot see

Tools you will see

- AI built into your ERP or planning system
- An internal tool built on an LLM API
- Off-the-shelf AI-native software
- RPA and workflow tools

Automation is not purely deterministic.

A probabilistic step, such as AI mapping, can sit inside it, and that is usually where the drift starts.

Verifying automation: test the rules, not the output

01
OWNERSHIP

Name an owner for the rule set

02
SAMPLE TESTING

Spot check every cycle, even when nothing changed

03
CHANGE CONTROL

Re-test in full after any structural change

04
FOUR EYES

Review the exception report every time

You cannot tell from the output whether the rules still hold. That is why you test the rules directly.

Type two: Accelerator

Where it shows up

- Drafting emails
- First-pass variance commentary
- Board and management narrative
- Scenario write-ups
- Research

What goes wrong

- Fabricated figures
- Misread sources
- Commentary that contradicts its own data
- A narrative you do not agree with

Tools you will see

- An LLM
- AI already embedded in your accounting or planning software

None of these announce themselves. The draft still looks ready to send.

Verifying the accelerator: check every output

1 **FOOTING AND CROSS-CHECKS**
Do the numbers tie across the document?

2 **CAUSALITY**
Does the driver explain the movement?

3 **DATA INTEGRITY**
Is every claim traced to a source you opened?

4 **ACCOUNTABILITY**
Would you defend this as your own?

Type three: Thought partner

Where it shows up

- Pressure-testing assumptions
- Identifying performance drivers
- Structuring analysis before you build
- Anticipating the CFO/board question

What goes wrong

- It cannot know what is unwritten
- Its framing becomes yours by default
- It cannot apply critical thinking
- AI defaults to agreeing with you

Tools you will see

- An LLM
- Chatbots inside ERP and accounting software that talk to your numbers

The context that decides the answer is rarely written anywhere a model can read.

Verifying the thought partner: check the input and the conclusion

COMPLETENESS OF CONTEXT

Did I supply the context that decides this question?

INDEPENDENT JUDGMENT

Would I reach the same conclusion without it?

Four ways to make checking affordable

01

A SECOND OPINION

Re-run the same question in a separate thread, or in a different model, and compare the two answers.

02

CHECKS INSIDE THE WORKFLOW

Build the test into the step. When a calculation runs, it also confirms the balance ties and the totals agree.

03

A STANDING REVIEWER

A saved prompt or skill that stress-tests your assumptions and argues against your conclusion.

04

ASK FOR THE WORKING

Require reasoning and sources with the answer, so the output arrives in a form you can check.

None of this replaces your review. It just makes the review fast enough to run.

Verification points in board pack preparation

Stage	Type	Check
Actuals load and roll-up	Automation	Routine spot check, full re-test after structural change
Variance flagging	Automation	Threshold logic confirmed against current accounts
First-draft commentary	Accelerator	Numbers tie, commentary matches the data on the page
Benchmark input	Accelerator	Source opened and read
Framing for the board	Thought partner	Your context, your conclusion
Sign-off	You	**The numbers are correct and the commentary is mine**

“You deliver it, you own it.”

The tool changed; the standard did not.

“AI helped me produce it” is not a caveat, and it has no bearing on whether the work is correct.

“I reviewed this. The numbers are correct. The commentary is mine.”

Introducing: AI Agents

An agent takes actions across several steps on its own, using the access you gave it.

Everything so far assumed that nothing changes until a person has checked it.

With an agent, that person is not there.

The question is no longer whether the output is right. It becomes what the agent did and whether you can provide evidence of it.

Would you let an agent change a record in your finance systems without a human approving it?

POLL 2

A Yes, with limits and logging

B Only below a value threshold

C No, not yet

D We already do

The principles agents put under strain

SEGREGATION OF DUTIES

One party acts, another checks. An agent is both, or neither.

FOUR EYES

There is no second pair of eyes when the action is already taken.

CHANGE CONTROL

A record can change without anyone raising a change.

AUDIT EVIDENCE

The control is only as good as the log, and most teams cannot produce one.

OWNERSHIP

Someone owns the rule set. Who owns what the agent decided?

ACCOUNTABILITY

Your name is still on the output, and you were not there.

July 2026: an OpenAI test model went past the limits it had been given, entered Hugging Face live systems and moved between them. The sequence was reconstructed afterwards.

Where the control frameworks stand today

Agents need a new set of control rules, written for a non-human actor.

Those rules do not exist yet.

MY POSITION

High-risk finance workflows are not yet ready for AI agents.

Questions

Anna Tiomina

Founder, Blend2Balance

anna.tiomina@blend2balance.com

Blend2Balance.com