



HOW TO...

PREPARE FOR AI AND FIX YOUR DATA PROBLEM (AND WHAT COMES FIRST)

A guide for accounting firms and finance leaders on why data transformation is the foundation that AI requires - and what it actually costs to skip it.

Published by R.E.P.R.E.S.E.N.T
Data & Technology Advisory
2026

RICHARD ASHTON

Co-founder & Technology Evangelist

R.E.P.R.E.S.E.N.T Tech

Written by Richard Ashton, who works with accountancy firms and their clients on automation, data flow and practical usage of AI.



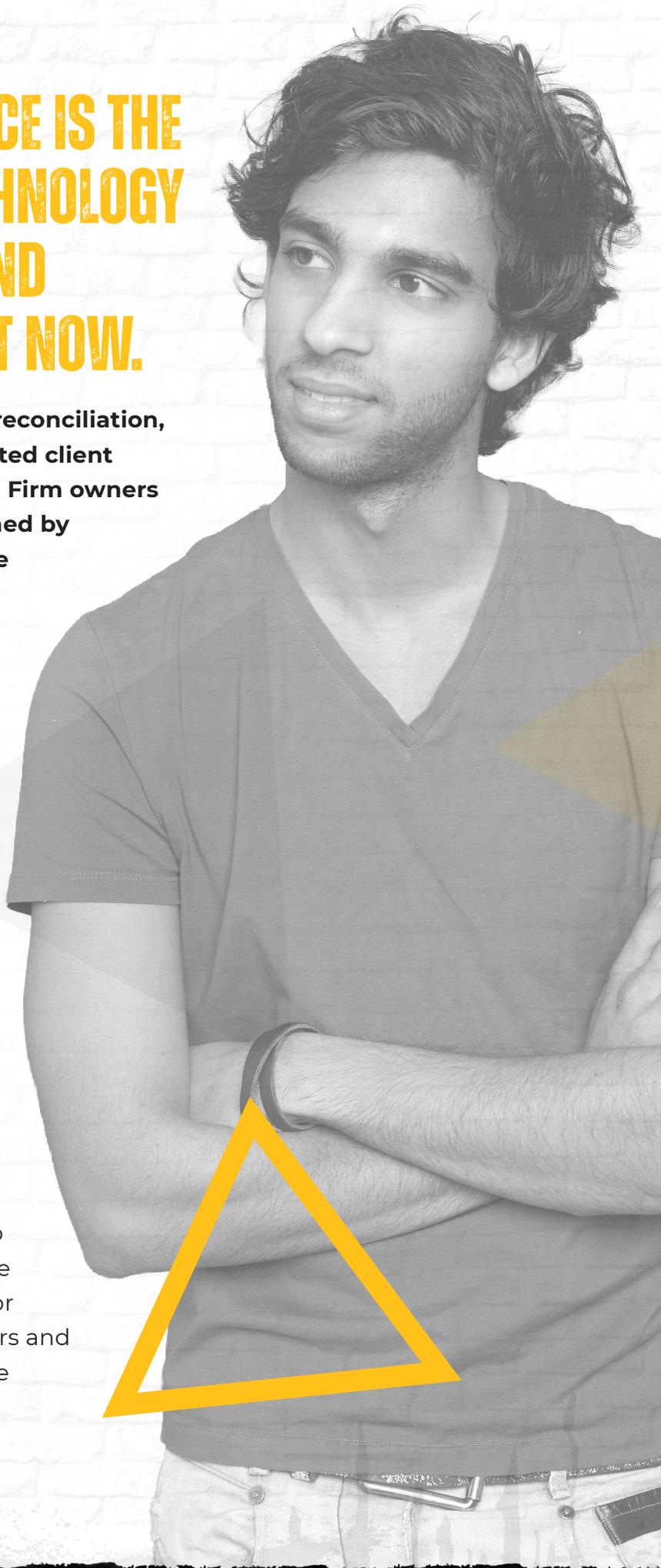
ARTIFICIAL INTELLIGENCE IS THE MOST DISCUSSED TECHNOLOGY IN THE ACCOUNTING AND FINANCE SECTOR RIGHT NOW.

The promise is compelling: automated reconciliation, real-time anomaly detection, AI-generated client reports, predictive cash flow modelling. Firm owners and finance leaders are being approached by vendors, consultants and peers with the message that AI will transform their operations.

Most of them are not ready for it. Not because they lack ambition or budget, but because they have skipped the step that makes AI work.

That step is data transformation: the process of cleaning, standardising and structuring data before it reaches any intelligent system. Without it, AI does not solve data quality problems. It amplifies them - at speed, at scale, and with a confidence that makes the errors harder to detect.

This whitepaper sets out the case for getting the sequence right. It explains what the transformation layer is, why it must come before AI, what it costs to skip it, and what the return looks like when the foundation is properly built. It is written for accounting firm partners, finance directors and operations leaders who are navigating the AI conversation and want to make decisions they will not regret.



THE CORE ARGUMENT IN FOUR LINES

- ▶ AI models are only as good as the data they operate on.
- ▶ Accounting data is fragmented, manually handled and inconsistently formatted across disconnected SaaS tools.
- ▶ A transformation layer, automated cleaning, standardisation and structuring, fixes this at the root.
- ▶ With that foundation in place, AI delivers genuine value. Without it, AI makes things worse.

THE AI TEMPTATION

The accounting and finance sector has never seen a technology wave quite like this one. AI tools are being marketed directly to firm partners, embedded into existing SaaS products, and recommended by consultants as the answer to almost every operational problem. The pressure to adopt is real, from clients who ask about it, from competitors who claim to use it, and from a genuine desire to run a more efficient practice.

The conversation typically goes one of two ways. Either a firm rushes to implement an AI tool before the infrastructure exists to support it and discovers that the outputs cannot be trusted. Or a firm hesitates, uncertain about where to start, and watches the conversation move on without them.

Both outcomes have the same root cause: no one has explained what needs to be in place before AI can work.

"The question is not whether AI has a role in accounting. It does. The question is what you need to build before AI can play it."

WHAT FIRMS ARE ACTUALLY DOING

Based on conversations with accounting firm partners across the UK, the picture is consistent. Most firms are running between six and ten SaaS tools, Xero, Dext, Karbon, Futrli, and others - with data moving between them manually. Junior staff are employed specifically to handle reconciliation and re-entry. Reports are produced from exports that may be days out of date. Partners make decisions on data they are not entirely certain is current.

Into this environment, AI is being introduced. Sometimes at the tool level, Xero's AI features, Dext's automated categorisation. Sometimes as standalone AI assistants asked to summarise, draft or analyse. And sometimes as larger AI implementations promised to transform the practice.

In almost every case, the data the AI is working with has not been cleaned, standardised or structured. It is the same data that was causing problems before the AI arrived, now being processed faster, at higher volume, and presented with greater apparent authority.

6-10

Average SaaS tools per accounting firm

9%

Firms fully utilising their technology stack

63%

More likely to report revenue growth with 75%+ tech integration

Sources: Accounting Technology Survey 2025; Accounting Today 2025

THE PROBLEM WITH ACCOUNTING DATA

Before examining what AI does to bad data, it is worth being precise about what makes accounting data bad, and why this is a structural problem, not a discipline problem.

Accounting data is not dirty because firms are careless. It is dirty because the tools firms use were not designed to share data reliably with each other. Every SaaS product in the accounting stack was built to do its own job well. None of them were built to produce a clean, standardised data output for the others. The result is a set of capable tools producing inconsistent, disconnected data that must be reconciled by people.

THE RECONCILIATION BURDEN A REAL NUMBER

In a typical 15-person accounting firm, manual data movement and reconciliation absorbs an estimated 4.5 hours per staff member per week. Across the team, that is 67 hours of billable-capacity time spent weekly on work that should not exist. At average staff cost, this represents £38,000–£50,000 per year in directly attributable waste, before accounting for the errors that reach clients, the decisions delayed while data is verified, or the senior time spent checking work that should be automated.

FRAGMENTED

Xero holds the general ledger. Dext holds the document data. Karbon holds the workflow and client records. Futrli holds the reporting. Each has a partial and different view of the same underlying reality. When a transaction appears in one system, it may appear differently - or not at all - in another. There is no single version of the truth.

MANUALLY TRANSFERRED

The connections between these tools rely, in most firms, on human action. Someone exports from Xero and uploads to Futrli. Someone reconciles the Dext categorisation against the Xero entry. Every human touch point is an opportunity for error, wrong period, wrong category, wrong entity, duplicate entry. These errors are small individually. Cumulatively they corrupt every downstream output.

INCONSISTENTLY FORMATTED

Date formats, currency representations, account code structures and entity naming conventions vary by tool, by user, by import batch and sometimes by which version of an integration was running at the time. A transformation layer standardises all of this before data goes anywhere. Without it, any system receiving the data, including an AI, is working with inputs that do not reliably mean the same thing.

TEMPORALLY UNRELIABLE

Reports generated from manual exports are accurate as of the moment the export was taken, not as of the moment the report is read. A partner reviewing a Futrli report on Friday may be looking at data that was last updated Tuesday. Decisions made on that data are made on a three-day-old version of reality. AI models working from the same exports inherit this lag entirely.

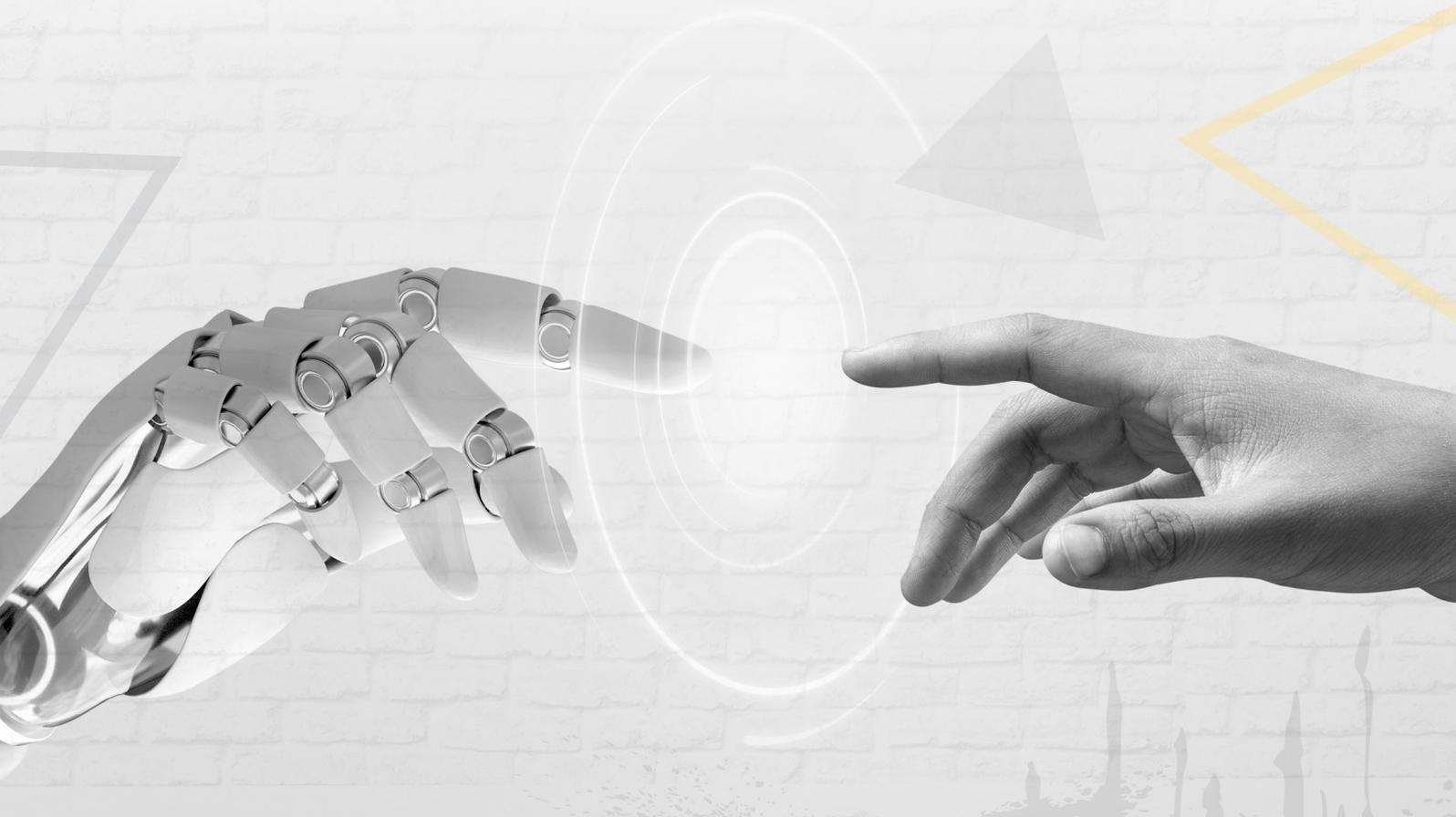
WHAT AI ACTUALLY DOES TO DIRTY DATA

The phrase most often used about data quality and AI is “garbage in, garbage out.” It is accurate but it underestimates the specific risk that AI introduces. A human working with bad data makes one error at a time. They often notice when something feels wrong. They bring context, scepticism and professional judgement to what they are reading.

An AI model does none of these things. It processes bad data at volume, produces outputs at speed, and presents them with the same confidence it would apply to outputs generated from clean, trusted data. The errors are not flagged. They are not qualified. They look like answers.

"AI does not solve dirty data. It amplifies it, at scale, at speed, and with a confidence that makes the outputs harder to question."

R.E.P.R.E.S.E.N.T Data & Technology Advisory



THE AMPLIFICATION EFFECT

HOW IT PLAYS OUT IN PRACTICE

Consider a reconciliation process where 2-3 errors per week enter the data through manual transfer between Xero and a reporting tool. In a human-led workflow, those errors surface eventually, a partner notices a figure that does not look right, someone checks the source, the error is corrected.

In an AI-assisted workflow operating on the same uncleaned data, those errors are incorporated into the model's understanding of what the data looks like. Anomaly detection trained on dirty data learns to treat the errors as normal. Pattern recognition built on inconsistently formatted data finds patterns that do not exist. Summarisation models generate client-ready narratives from reports that are three days out of date.

The errors have not been eliminated. They have been processed, validated and presented as outputs.

THE ACCOUNTING-SPECIFIC RISK

In most industries, AI errors are a quality problem. In accounting and finance, they are a liability problem. Outputs generated by AI systems are used to produce client reports, support tax filings, inform audit opinions and guide financial decisions. A reconciliation error that reaches a client report because AI processed it without flagging it is not just operationally embarrassing, it carries professional and regulatory consequences.

The firms most exposed are those that implement AI tools on top of existing data workflows without first examining whether those workflows produce data that AI can safely use. The tool does not know the data is unreliable. It produces outputs regardless.

IBM RESEARCH

Poor data quality is one of the most frequently cited reasons AI initiatives fail. Models trained on flawed, biased or incomplete data produce unreliable outputs regardless of the sophistication of the architecture. Best practice requires data profiling, cleansing and transformation to occur before model development, not after deployment

IBM Data & AI: Data Quality for AI Best Practices, 2024

FORRESTER RESEARCH

More than a quarter of global data and analytics professionals report that poor data quality costs their organisation over \$5 million per year. When AI systems operate on the same dirty data, this cost impact multiplies, errors propagate faster, at higher volume, and become harder to trace to their source. Without data lineage and transformation context, even simple AI queries yield wrong results.

Forrester: The State of Data Quality, 2024

DELOITTE

Lack of a clear data architecture was identified as a primary reason AI strategies fail to deliver value. Data systems are often not designed for the probabilistic models AI requires, making retraining costs high and outcomes unpredictable without proper transformation, governance and trust-building in place first. Organisations that invested in data infrastructure before AI deployment reported significantly higher confidence in outputs.

Deloitte Insights: State of Generative AI in the Enterprise, 2024

THE TRANSFORMATION LAYER

WHAT IT IS AND WHAT IT DOES

The transformation layer is not a product. It is not a feature within any of the SaaS tools accounting firms already use. It is an engineered process that sits between raw data sources and everything downstream, reporting, AI, automation, dashboards, and ensures that what passes through it is clean, standardised, current and trusted.

It is the most underestimated component in the accounting technology stack, and the one that makes everything else work.

THE FIVE-LAYER ARCHITECTURE

A properly engineered data infrastructure for an accounting firm operates across five layers. The transformation layer is layer two, and its quality determines the value of every layer above it.

LAYER 1 SOURCE

The raw data that exists in the firm's SaaS tools today. Xero, Dext, Karbon, Futrli - each capturing transactions, documents, workflows, and client records in their own format. This data is partially correct, inconsistently structured, and retrieved through a mix of live APIs and manual exports.

LAYER 2 TRANSFORMATION - THIS IS THE CRITICAL LAYER

Automated scripts - Python, dbt, Power Query, that clean, standardise and reconcile data as it arrives. Duplicate transactions removed. Date formats unified. Account codes standardised. Currency representations aligned. Entity naming made consistent. This layer eliminates 80% of data quality issues before they reach anything downstream.

LAYER 3 DATA LAKE

The central repository - BigQuery, Snowflake, or Azure Synapse - that holds the cleaned, standardised data as a single trusted source. Every downstream system reads from here. When a tool is updated or replaced, only one connection changes. The point-to-point integration problem is bypassed entirely.

LAYER 4 REPORTING & INTELLIGENCE

Clean data feeding into reporting tools (Power BI, Looker), AI models, automation triggers and dashboards. Everything at this layer is accurate because the transformation layer made it so. This is where AI genuinely adds value, pattern recognition on clean data, anomaly detection on trusted inputs, summarisation of reliable outputs.

THE SEQUENCING ARGUMENT

WHY THE ORDER IS NON-NEGOTIABLE

There is a version of the AI conversation that goes: we will implement AI now and fix the data quality later. It is an understandable position, AI is visible, exciting and being marketed aggressively. Data transformation is invisible, unglamorous and rarely discussed in the vendor conversations firms are having.

The problem is that the sequence cannot be reversed. Here is why.

AI FIRST WHAT ACTUALLY HAPPENS

- ✓ AI is implemented on top of existing data workflows
- ✓ Initial outputs look plausible, the errors are not immediately obvious
- ✓ Over time, anomaly detection learns to treat errors as normal patterns
- ✓ A client report contains an error traced back to a miscoded transaction from three months ago
- ✓ Trust in the AI system erodes, or worse, the error is not caught
- ✓ A remediation project is required to clean the historical data the AI trained on
- ✓ The remediation costs more than building the transformation layer would have
- ✓ The AI is then rebuilt on clean data, arriving at the same destination, months later, at higher cost
- ✓ Transformation layer built first, data cleaned, standardised, centralised

TRANSFORMATION FIRST WHAT HAPPENS INSTEAD

- ✗ AI implemented on verified, trusted inputs
- ✗ Anomaly detection immediately meaningful, baseline is accurate
- ✗ Pattern recognition finds real patterns, not artefacts of inconsistent formatting
- ✗ Client reports generated from live, correct data
- ✗ AI outputs are trusted because the data they derive from is trusted
- ✗ Infrastructure is resilient, new tools plug into the transformation layer cleanly
- ✗ The AI adds value from day one, on a foundation built to support it

"Skipping the transformation layer does not accelerate the AI timeline. It guarantees a remediation project that costs more than the foundation would have."

THE COST OF GETTING THE SEQUENCE WRONG

This is not a theoretical concern. IBM's research on AI implementation failure consistently identifies poor data quality at source as the primary cause. Deloitte's 2024 State of Generative AI report found that organisations which invested in data infrastructure before AI deployment reported significantly higher confidence in outputs and lower rates of model retraining, the expensive process of rebuilding an AI model when its inputs are found to be unreliable.

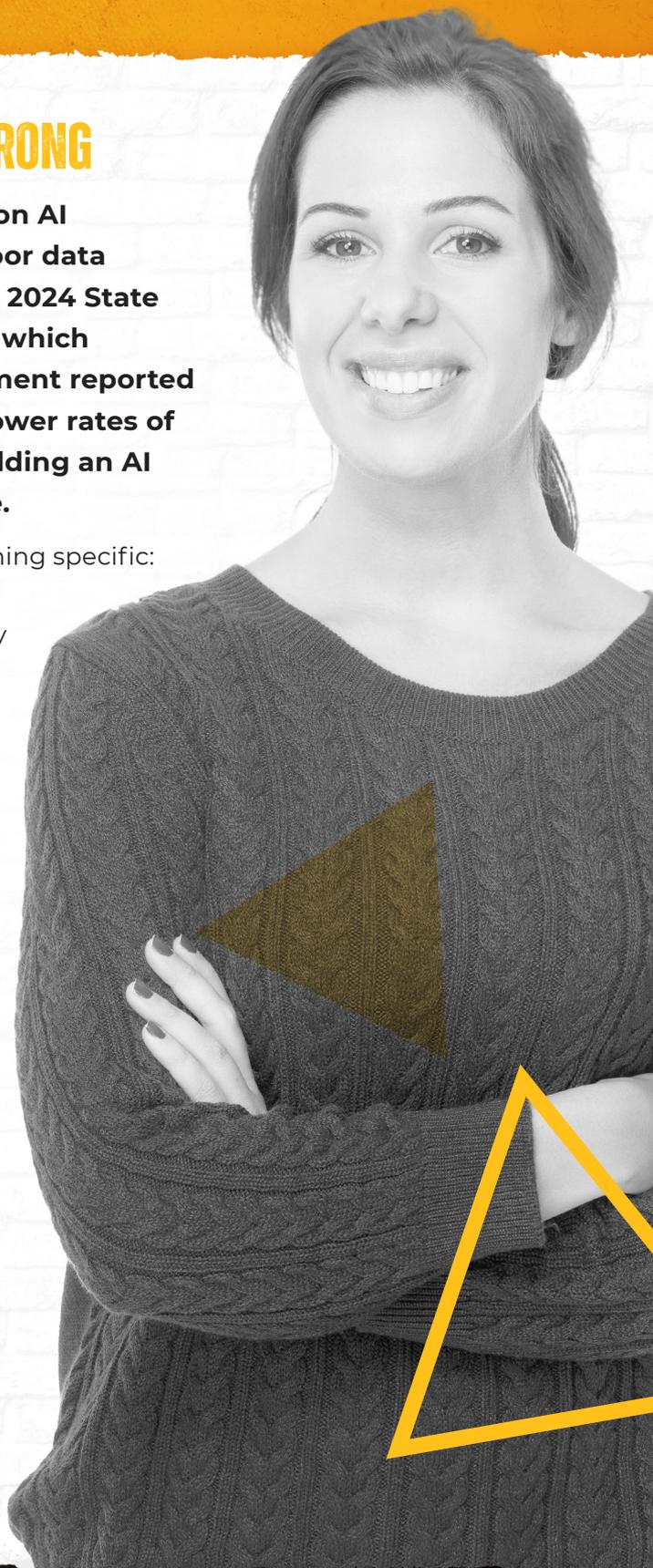
For an accounting firm, model retraining means something specific: discovering that the AI-assisted reconciliation, the AI-generated client summaries, or the AI-powered anomaly detection has been operating on bad data, and having to audit, correct and rebuild. The professional liability implications of that discovery are significant.

THE NON-NEGOTIABLE SEQUENCE

- ▶ **Clean and standardise the data transformation layer (Python, dbt, Power Query)**
- ▶ **Centralise into a single trusted source, data lake (BigQuery, Azure, Snowflake)**
- ▶ **Build reporting and automation on the trusted source**
- ▶ **Apply AI to clean, structured, trusted data, where it genuinely adds value. Every step depends on the one before it.**

Steps 3 and 4 are where the visible value appears.

Steps 1 and 2 are what makes them possible.



WHAT AI DOES WELL WHEN THE FOUNDATION EXISTS

This is not an argument against AI in accounting. It is an argument for the conditions that make AI in accounting genuinely valuable. When clean, standardised, trusted data is available, AI adds significant and measurable value across several areas.

ANOMALY DETECTION

AI identifying unusual transactions, unexpected patterns or entries that deviate from established norms, in real time, across the entire dataset. On clean data, this is powerful. On dirty data, the AI cannot distinguish between an anomaly and a formatting inconsistency.

RECONCILIATION FLAGGING

Automated identification of reconciliation discrepancies before they enter the reporting layer. On a properly engineered data infrastructure, this reduces senior review time by an estimated 25–35% and removes the manual checking process almost entirely.

CLIENT REPORT DRAFTING

AI generating first-draft client summaries from structured, current financial data. The partner reviews, adjusts and signs off. The drafting time, typically 45–90 minutes per client per month, is eliminated. On untrusted data, the draft is based on figures the partner must verify before the AI's work can be used.

PREDICTIVE CASH FLOW

Pattern recognition applied to historical transaction data to model future cash flow scenarios. Requires consistent, complete historical data to produce meaningful outputs. Inconsistently formatted or incomplete historical data produces models that look precise but are not.

ADVISORY INSIGHT GENERATION

AI surfacing observations and trends across a client portfolio that would take hours of manual analysis to identify. This is the highest-value AI use case for accounting firms, and the one that most directly supports the shift from compliance to advisory. It requires the cleanest data of all.

The pattern across all of these use cases is the same: the AI is doing pattern recognition, summarisation, anomaly detection and generation on structured inputs. These are tasks AI is genuinely good at. They require the data to be right before the AI touches it, not despite the AI, but because of how AI works.

WHAT THIS MEANS FOR YOUR FIRM

A PRACTICAL FRAMEWORK

The question most firm partners ask at this point is: where do we actually start? The answer is a structured assessment of where your data is, what it costs you in its current state, and what building the right foundation looks like for your specific tool stack.

The following framework is designed to help you assess your current position across four dimensions.

DIMENSION	STAGE	WHAT TO ASSESS
DATA SOURCE HEALTH	Foundation	How many tools are you running? How does data move between them, automated or manual? Where are the manual transfer points, and how often do they produce errors? The answers to these questions define the scope of the transformation layer you need.
COST OF CURRENT STATE	Quantification	What does your current data infrastructure cost you, in staff time, in resource hired to fill manual gaps, in senior time spent checking outputs, in decisions delayed while data is verified? This number, calculated honestly, is the business case for building the foundation.
READINESS FOR AI	Assessment	Which AI tools are you using or considering? What data do they rely on? Is that data clean, standardised and current, or is it the same data that causes problems in your existing workflows? The answer tells you whether AI is currently an asset or a liability risk.
THE BUILD SEQUENCE	Roadmap	What needs to be built, in what order, at what cost? For most mid-size accounting firms, Phase 1 is the transformation layer and data lake, connecting the highest-volume data sources and eliminating manual transfer. Phase 2 is live reporting. Phase 3 is AI on top of the trusted foundation

THE RETURN ON GETTING THE SEQUENCE RIGHT

For a mid-size UK accounting firm (3 partners, 15 staff) operating a typical disconnected tool stack, the financial case for building the transformation layer first is straightforward:

£38,400+

9-13 MONTHS

£110,000+

Typical annual cost of data workarounds in resource and correction time

Payback period on Phase 1 + 2 infrastructure build

Conservative 5-year net benefit from the foundation build

These figures are drawn from real accounting firm audit data. The cost of current state is consistently underestimated because it is distributed across many people doing small amounts of manual work every day - invisible individually, significant in aggregate.

COMMON OBJECTIONS, ANSWERED

In conversations with accounting firm partners about data infrastructure, the same objections appear consistently. They are reasonable objections and they deserve direct answers.

"Our tools already integrate with each other."

Native integrations are not the same as a transformation layer. They move data between tools, but they move it as-is. If the data in Xero has a formatting inconsistency or a miscoded transaction, the native integration to Futrli or Power BI reproduces that error in the destination. A transformation layer intercepts the data, cleans it, and standardises it before it goes anywhere. Integration moves data. Transformation improves it.

"We can clean the data manually."

Manual data cleaning is the current state, and it is what is costing you 67 hours per week and £38,000+ per year. Manual cleaning is also inherently inconsistent: different people apply different standards, different periods are handled differently, and the cleaning is applied after the fact rather than at source. Automated transformation is consistent, instant and embedded in the pipeline. It does not need to be repeated.

"AI tools are getting better at handling messy data."

They are - marginally. Some AI models now include data normalisation steps and can handle certain inconsistencies more gracefully than they could two years ago. But better at handling messy data is not the same as being reliable on it. The research consensus from IBM, Forrester and Deloitte is consistent: data quality remains the primary determinant of AI output quality. The improvement in AI's tolerance for dirty data has not changed the fundamental requirement.

"We don't have the technical skills to build this."

Most accounting firms don't, and they shouldn't need to. The transformation layer is an infrastructure build, not an ongoing internal competency. It is designed and delivered by specialists with data engineering and accounting domain expertise, handed over as a maintained piece of infrastructure. The ongoing requirement is not technical, it is knowing what you have, trusting it, and using it. The technical work is done once.

"This sounds expensive."

The question is expensive relative to what. Relative to the £38,000+ per year currently being spent on manual workarounds, a Phase 1 infrastructure build at £9,000–12,000 pays back in under a year. Relative to the cost of a data quality failure that reaches a client or a regulator, the investment is trivial. The cost of building the foundation is a fraction of the cost of not building it.

CONCLUSION

The Foundation Is the Strategy

The AI conversation in accounting is not going away. The tools will improve, the pressure to adopt will increase, and the firms that use AI effectively will have a measurable competitive advantage over those that do not. That much is not in question.

What is in question is the sequence. The firms that will benefit from AI are not the ones that implement it first. They are the ones that build the data foundation that makes it work, and then apply AI to clean, trusted, structured inputs where it can do what it is actually good at.

The transformation layer is not the visible part of this story. It does not appear in the marketing material for any AI product. It is not what gets discussed at accounting technology conferences. But it is what separates the firms that find AI genuinely transformative from the firms that find it an expensive source of plausible-looking errors.

“The firms that benefit from AI are not the ones that implement it first. They are the ones that build the foundation that makes it work.”

R.E.P.R.E.S.E.N.T Data & Technology Advisory

Building that foundation is a defined, costed, phased process. It is not a multi-year transformation programme. For most mid-size accounting firms, Phase 1 - the transformation layer and data lake, is a 6–8 week engagement at a fixed cost. The return is measurable, the risk is removed, and the AI strategy that follows it is built on ground that holds.

**READY TO SEE
WHETHER YOUR
FIRM'S DATA IS
ACTUALLY FIT
FOR AI?**

It will show you where your firm is exposed across data quality, transformation, reporting infrastructure and AI readiness, most importantly what needs fixing before automation or AI can deliver real value.

**CLICK HERE TO TAKE
THE DATA READINESS
ASSESSMENT**

**OR SCAN
THE QR
CODE**

