Getting data in order before AI

For COOs and heads of operations and data at insurers, platforms, wealth managers and advice firms: why AI tools that work in a demonstration struggle on a firm's own records, and four questions to settle about the data before the project starts.

IN BRIEF

The problem

An AI tool that performs well in a demonstration meets the firm's own records and produces more exceptions than planned, so people end up checking far more than planned.

Why it happens

Years of inconsistent fields, free text, duplicates and scanned documents sit in the systems an AI tool has to read, and they were not measured first.

What to do

Measure the quality of the data the tool will use, decide who owns it, and budget for fixing it before the tool is bought.

01

Why AI projects stall on data

A demonstration runs on clean data; a firm's own records are rarely clean. Policy and client data built up over decades, across acquisitions and system changes, carries inconsistent fields, free text where structured fields should be, duplicates, gaps and documents stored as images. An AI tool reading those records makes more mistakes, flags more exceptions and needs more checking than the business case assumed.

Firms know this is a risk. In the Bank of England and FCA survey published in November 2024, four of the five risks firms rated highest for their current use of AI were about data: privacy and protection, quality, security, and bias and representativeness.

02

What poor data does to the business case

The cost shows up in exceptions: records the tool can't process cleanly, which people then handle by hand. As the exception rate rises, the saving shrinks, and at some point it disappears.

ILLUSTRATIVE EXAMPLE

The same tool on cleaner and messier data

Records the tool can't process cleanly	Hours of manual handling a month	Saving a month after the tool's cost
2%	50	£49,667
10%	250	£41,667
30%	750	£21,667
50%	1,250	£1,667

Illustrative: 10,000 records a month; fully manual processing at 10 minutes a record costs £66,667 a month at £40 an hour; the tool costs £15,000 a month; each exception takes 15 minutes to handle, and records the tool processes cleanly need no further handling.

The tool is the same in every row. The difference is the data it reads. A business case that assumes the first row, on data that behaves like the third, will promise more than twice the saving it delivers.

03

Four questions before the project starts

1. Which data will the tool use, and who owns it? Name the systems, fields and documents, and the person accountable for each one's quality. Data without an owner tends to stay as it is.

2. How good is it, measured rather than assumed? Check completeness, consistency and timeliness on a sample before any tool is chosen. The measurement takes days; finding out after go-live can take months.

3. Can you trace it? Know where each important field comes from, what changes it, and which version is right when two systems disagree.

4. Are you allowed to use it this way? Check the lawful basis, and whether using the data this way fits what customers were told when it was collected.

> FROM EXPERIENCE
>
> At Aviva I carried out unit-linked asset and liability reconciliations. At Prudential my work included valuation processes for Sarbanes-Oxley (SOx) compliance.
>
> Both depended on the same thing: knowing what the data meant, where it came from and where it was wrong. That work is slower than the technology, and it decides whether the technology works.

04

Fix first, or fix as you go?

Not every data problem has to be solved before an AI project starts. Fix first where errors would reach customers or regulatory reports, or where the exception rate

would wipe out the saving. Fix as you go where the tool itself can help: AI can find duplicates or extract fields from old documents, under close checking, as a first step in cleaning the data. Either way, put the data work in the business case and the budget, not in a footnote.

Questions for the board

01 Which data will each AI tool rely on, and who owns its quality?

02 Have we measured the quality of that data, or assumed it?

03 Can we trace where each important field comes from and what changes it?

04 What exception rate does the business case assume, and what happens to the saving if it doubles?

05 Do we have the lawful basis, and our customers' trust, to use this data in this way?

06 What would it cost to fix the data first, and is that in the budget?

Data quality decides whether an AI tool saves money or creates checking work. Measuring it before the project starts costs little, and it makes the business case honest.

If this is on your table, a 30-minute conversation is usually enough to tell whether I can help.

Talk to Mark mark@markwoodruff.co.uk · 07391 820971 · markwoodruff.co.uk

ABOUT THE AUTHOR

Mark Woodruff is a Chartered Actuary (FIA) and Chartered Financial Planner (FPFS). He spent 15 years pricing and building products at Aviva, Prudential, Partnership and St. James's Place, then more than seven years founding and running advice firms through to a substantial sale. Through Mark Woodruff Consultancy he works with financial services firms on product, pricing, distribution and governance.

Regulatory notice. Mark Woodruff Consultancy works with businesses only. I don't give regulated personal financial advice to individuals.

Disclaimer. This briefing is general information for financial services firms. It is not tailored to any firm's circumstances and is not legal, regulatory or financial advice. Work is agreed in a written letter of engagement before it starts.

Published October 2026
Regulatory references correct at October 2026
© Mark Woodruff Consultancy 2026