peter brady

FINDINGS & RECOMMENDATIONS

Private AI Assistant Discovery for [redacted]

A review of how [redacted] works with information today: where it lives, what your team is already doing with public AI, the first thing worth building into a private AI assistant, and the roadmap for widening it across your other systems.

PREPARED FOR	PREPARED BY
[redacted], a UK [redacted] charity	Peter Brady
MAIN CONTACT	STATUS · REFERENCE
[redacted]	Confidential · DISC-[redacted]

In short

Your information is spread across four main systems that do not talk to each other, so answering an ordinary question, how a fund is doing, who has not been thanked, what a case has cost, means someone exporting spreadsheets and stitching them together by hand.

In the meantime, several staff are already pasting confidential material into public AI tools to get through the work faster, which is understandable and also a real exposure.

There is a clean first step. One AI assistant, running inside your own systems, that answers the handful of questions your team asks every week, with every answer linking back to the record it came from. It gives people the speed they are currently going to ChatGPT for, without the data ever leaving your control.

4	3 of 4	9
core systems holding the same answers in pieces	shadow-AI signals present across the team	recurring questions worth answering directly

01 What this review covered

This was a fixed-price discovery, not a build.

You cannot build anything worth having until three things are true: you know where the answers actually live, you know what your team is already doing while they wait for them, and you agree what a right answer means. Get one of those wrong and you build the wrong thing well, which is the expensive way to fail.

So I looked at three things, and this report is the output. There is no obligation to go further.

- **Where your data actually lives.** The systems in daily use, what each one holds, who looks after it, and how joined-up it really is.
- **What your team is doing with AI today.** Not by asking them to confess on a form, but by looking at the signals the work itself leaves behind.
- **The questions worth answering.** Drawn from your real data with you, then pinned down so we agree exactly what each one means before anything is built.

Those three are the rest of this report, in that order.

02 Where your data lives today

The same facts about a supporter, a fund or a case are held in more than one place, in more than one shape. Nothing here is unusual for a charity your size. It is the reason ordinary questions take half a day.

The measure that matters is not whether each system is well kept. Two of yours plainly are. It is whether any one of them can answer a whole question on its own, and none of them can.

SYSTEM	WHAT IT HOLDS	LOOKED AFTER BY	STATE
Salesforce (NPSP)	Supporters, donations, memberships, appeals	[redacted]	Well kept
Xero	Accounts, restricted funds, banking	[redacted]	Well kept
Google Sheets	Volunteer hours, field counts, event sign-ups	[redacted]	Patchy
SharePoint & Drive	Policies, case notes, funder contracts, minutes	[redacted]	Unmanaged

The pattern is the usual one: the CRM and the accounts are each tidy on their own, but no single place is authoritative when a question crosses them. A fund's total sits in Salesforce as gifts and in Xero as banked income, and the two rarely match to the penny, for good reasons that no one currently has time to reconcile.

That gap is not only an inconvenience. Work still has to go out on Friday, so people find their own way round it, and that is the next thing this report looked at.

03 What your team is already doing with AI

Asking a team what AI tools they use gets you a clean survey and a false picture. People know that pasting a case note into a public chatbot is off-side, so they do not tick the

box. I work the other end: what the output shows. Three of the four signals I look for are present here.

- **The house style shifts mid-document.** Several recent bids and reports carry the tell-tale evenness of a public model, then drop back into your own voice. It reads like work drafted elsewhere and pasted back.
- **Confidential material is leaving the building.** At least [redacted] staff described, once it was safe to, pasting real case details or supporter information into ChatGPT to summarise or reword. This is the exposure that matters most.
- **The same task is being redone by hand.** The monthly fund figures are rebuilt from scratch each time because no one trusts last month's spreadsheet. This is exactly the work an AI assistant should carry.

The fourth signal, a paid team AI account already in use, is not present. Nobody has bought their way to safety yet, which means the exposure is entirely on personal, public tools.

None of this is a discipline problem. Your team is reaching for the fastest tool that works. The answer is not a policy telling them to stop, it is a private AI assistant that reaches into your actual data systems, which is what makes it genuinely faster than a public chatbot, and it means the sensitive work finally has somewhere safe to go.

04 The questions worth answering

Working from your live data together, we drew out nine questions the team asks on a regular basis and currently answers by hand. The important part is not the list, it is that we agreed precisely what each one means. Definitions are where these projects usually go wrong, so we settled them first.

> **An active member** is someone whose membership has not lapsed as at the question date, counting the [redacted] grace period you apply in practice but had never written down.
>
> **A fund's total** is completed gifts recorded against that appeal source, before Gift Aid, matching how [redacted] reports it to the board.
>
> **Thanked** means a thank-you activity logged against the gift within [redacted], not simply that a letter template exists.

With those agreed, questions like "how is the [redacted] doing and who still needs thanking" or "what has this case cost us in staff time" become answerable in one step instead of an afternoon.

These definitions do more than steer the build. They are what the assistant gets marked against later. Each agreed question becomes a test with a known right answer, worked out from your records separately from the assistant, and the build is not finished until it returns that answer every time.

The [redacted] grace period is the clearest example of why this matters. It is a rule your team applies without thinking and had never written down, and an assistant that did not know it would return a member count that looked entirely reasonable and was quietly wrong. Mistakes of that kind are not the AI inventing things. They are it reading your world the way an outsider would. Pinning the rule down now is what makes it catchable later.

05 How the answers are checked

The fair worry about AI is that it makes things up. So the AI assistant is built so that where a number matters, it cannot. The figures are worked out in ordinary code against your live records, not guessed by the model, and every answer carries a link straight to the record it came from, so anyone can check it in a click.

Nothing has been built yet, so the two below are worked examples rather than results. The questions are yours, from the list above. The figures in them are invented. The shape of the answer, the reconciliation, and the link back to the record behind every claim are exactly what you would get.

How is the [redacted] doing, and does it match the accounts?

The appeal has raised £[redacted] across [redacted] **gifts** in Salesforce. Xero shows £[redacted] banked against the restricted fund as at the period close. The £[redacted] difference is [redacted] cheques recorded before the close but banked after it, listed in full. The books are right; the timing explains the gap.

Source: Salesforce appeal report · Xero fund P&L • FIGURES FROM CODE, NOT THE MODEL

What has [redacted] case involved, and what is still outstanding?

The case has [redacted] **logged notes** since [redacted], summarised faithfully: an initial home visit, two support calls, a referral, and a review. **One action is still open**: a follow-up call agreed at the last review and not yet marked done. Nothing in the notes is invented; each point links back to the entry it came from.

Source: CRM case record • SUMMARISED FROM THE RECORD ONLY

Worked examples are not proof, and I will not pretend otherwise. They show you what to expect, not that it works.

The proof comes later, and it will not be my word for it. Before the build goes live, every agreed question has to come back right, with working links to the source, 20 times in a row, marked against a figure worked out from your records separately from the assistant. One wrong answer out of the 20 fails the question, and it does not ship until it passes. That report is part of what I hand over.

The full method, including the mistakes this testing has caught on my own demonstration build and what I did about each one, is published openly at peterbrady.co.uk/do-private-ai-assistants-make-things-up

06 What I would build first

Not everything at once. The right first build is the one that removes the most weekly pain for the least risk, and proves the approach on ground you can check. That is the **fund and supporter AI assistant**: it reads Salesforce and Xero, answers the money and thank-you questions your fundraising and finance teams ask every week, and reconciles the two systems the way you do it by hand today.

It is one AI assistant, over two systems you already trust, doing a job with a clear right answer. Once it has earned that trust, widening it to the case notes and the volunteer figures is a small step, not a fresh project.

You might fairly ask why the first build is not the case notes, given that is where the exposure is worst. It is because trust has to be earned somewhere it can be checked.

A fund total is either right or wrong, and you can hold it against your own accounts in a minute. A case summary is a judgement, and you would be taking my word for it. So the first build goes where you can mark my homework, and the sensitive work follows once the method has proved itself on ground you can see.

07 Scope, price and what happens next

THE BUILD **Fund & supporter AI assistant** Salesforce + Xero, nine agreed questions, answers linked to source	RUNS **Inside your own tenancy** Your data never leaves your control; no public AI in the loop	PRICE **Fixed, £[redacted]** Quoted in full, staged across delivery, no day-rate creep
TIMELINE **[redacted] weeks** Built against the agreed questions, and not handed over until it is certified	CERTIFIED MEANS **Every question right, 20 times in a row** Each answer marked against a figure worked out separately from the assistant; one wrong answer in 20 fails the question	OPTIONAL WIDENING **Programme & volunteer figures** Google Sheets joins the same assistant, so field counts and hours answer alongside the money
OPTIONAL WIDENING **Case notes & documents** SharePoint and Drive, summarised faithfully, with a link back to every source	OPTIONAL WIDENING **More of your questions** The nine agreed questions grow as the team finds the next ones worth answering	ADDED ONLY WHEN IT LANDS **Once the first version has proved itself** Each widening scoped and quoted separately, on the same foundation, no commitment now

Built to outlast me

A fair question about any independent builder is what happens if I am not around. It is built on open standards and open-weight models, with no proprietary black box and no lock-in to me.

From there you choose: I can hand it over fully documented, so your own developers or another team can pick it up and run it, or I can support and maintain it for you on an ongoing retainer. Either way, you own it.

If you do nothing

The manual reporting carries on eating a day here and there, and the confidential material keeps going into public tools. Neither is a crisis today. Both are the kind of thing that becomes one on the wrong afternoon.

A Notes on your data

An honest caveat, because it matters. An AI assistant is only as right as the records under it, and a few of yours need a light touch before they can be trusted blindly: [redacted] are entered inconsistently across two fields, and a handful of older gifts are missing an appeal source.

None of this blocks the first build. It is handled in two ways: the AI assistant is pointed at the fields we know are clean, and where a record is ambiguous it says so rather than guessing. Tidying the rest is worth doing, and cheaper than it looks, but it is a choice for later, not a prerequisite.

That last point is worth drawing out, because it is a working part of the assistant rather than a caveat. Saying so rather than guessing is not politeness, it is checked on every answer.

Once the assistant has pulled figures from your systems and written its sentence, that sentence is checked back against what the systems actually returned before it reaches anyone: does it carry the figures the question turned on, does it claim anything the records do not support, is there a real source behind every claim? If not, the answer is thrown out and written again, and if it still does not hold up, the assistant tells you it cannot answer.

The testing proves the method before it reaches you. This runs on the answer you are actually given.