BURAK GALİBA
thoughts on AI

ESSAY · ENTERPRISE AI

Your Documents Are Not AI-Ready — and Retrieval Will Not Fix That

Twenty years of documentation is inventory, not a corpus. Why retrieval blends contradictions a human would catch — and the ten-question test that proves it.

DATA AI STRATEGY PILOTS

Burak Galiba · August 4, 2026 · 6 min read
https://burakgaliba.com/blog/documents-are-not-ai-ready

SCAN TO READ ONLINE

"We have twenty years of documentation" gets offered as an asset in almost every AI planning conversation I've been part of. Usually with some relief — the hard part sounds done. The knowledge exists. It's written down. Now we just point something at it.

It's inventory. It isn't a corpus. The distance between those two words is most of the project.

The demo already told you this, quietly

The prototype that worked ran on an export. Somebody chose which documents to include, noticed the two that contradicted each other, dropped the outdated one, and quietly answered a dozen small questions about precedence on the way to the spreadsheet.

That curation took an afternoon and nobody logged it. It was also the entire difference between a system that answered well and one that didn't — which is why a demo on frozen data proves so much less than it appears to. The judgement got applied by a person, once, and then deleted along with the export.

Reaching your documents is not the same as reading them

I've written before about the access layer — the reason most enterprise agents have nothing to connect to, and why the fix is a thin governed interface rather than a database login.

Solve that and you meet the second problem immediately. The agent can now reach the document store. That tells you nothing about whether the documents can answer anything.

Because enterprise documents were written for people who already know things. That assumption is invisible in the text and it is everywhere:

They omit the context the reader already has. A procedure that says "escalate to the regional team" is complete for the person who knows which region they're in, which team that maps to this quarter, and that the instruction hasn't applied to their unit since the reorganisation. None of that is in the document. It was never supposed to be.

They contradict each other without a precedence rule. Four documents give three answers. People resolve this constantly and mostly unconsciously — the one from the team that owns the process wins, unless it's older than the compliance note, in which case it doesn't. That ranking is real, it is load-bearing, and it exists nowhere except in the heads of people who have been there a while.

Their dates describe the file, not the content. "Last modified" records the day someone fixed a broken link. A document touched last week can carry a policy that stopped being true two years ago, and it will look fresher than the correct document nobody has opened since.

Nothing is ever deleted, only added. Superseded procedures don't get removed; they get quietly stopped being used. They remain perfectly retrievable, indistinguishable from current ones, and often better written — because the person who wrote the replacement was in a hurry.

Permissions live at the folder, meaning lives in the sentence. A document restricted to one group can contain a paragraph that is fine for everyone and a line that is not. Chunk it for retrieval and that folder-level boundary stops describing anything.

Humans detect contradictions. Retrieval blends them.

This is the part I'd put in front of anyone about to fund a knowledge assistant.

Hand a person four documents that disagree and they notice. The disagreement is the most interesting thing on the desk — they go find out which one is right, or they escalate, or at minimum they hedge when they answer.

Hand the same four to a retrieval system and it produces one fluent answer with no seam in it. The contradiction doesn't surface as a conflict. It surfaces as confidence. Whatever the model resolved to — the most repeated phrasing, the chunk that scored best, the passage that happened to be longest — arrives in the same tone as a fact everyone agrees on.

That is the failure mode that matters, and it is invisible by construction. A system that can't find an answer is annoying and self-reporting; people stop using it and someone raises it. A system that confidently averages four sources into an answer that was true in 2023 gets believed, and the trust it burns is inherited by whatever you build next.

The second half of the question is lineage

Quality asks whether the content is right. Lineage asks where it came from, who owns it, when it was last verified, and what happens to your system when it changes.

Lineage is the part that gets skipped, because at the moment of building it looks like paperwork. It stops looking like paperwork the first time an answer is wrong and the room wants to know why. Without lineage you cannot say which document produced it, whether that document was current, or who should have updated it. You can only say the model said it.

Gartner has been making a version of this argument for a while — its 2024 forecast that at least 30% of generative AI projects would be abandoned after proof of concept lists poor data quality first among four causes, alongside inadequate risk controls, escalating costs and unclear business value. That forecast window has now closed, so treat it as what it was — a prediction, not a measurement. The reason to take it seriously isn't the number. It's that data quality sits at the top of a list assembled before most of these projects had started.

The ten-question test

Before scoping any knowledge use case, spend an afternoon on this. It requires no tooling and no budget.

Take ten questions people actually ask — real ones, from real tickets or a real inbox, not ten you invented to be answerable. Find the answer in your own corpus, by hand, and record four things for each:

1. Was there a single unambiguous answer, or did you have to choose between sources?
2. If you chose, what rule did you use — and is that rule written down anywhere?
3. Did the answer depend on context that isn't in any document?
4. Was the best-looking source actually the current one?

The score is not the point. The pattern is. If you needed tacit precedence rules on six of ten, retrieval will need them too, and it doesn't have them. If three answers depended on knowing which region you're in, that's not a data-cleaning task — it's a scoping decision about what the system is allowed to be asked.

This is the same artifact I've argued a business unit should build instead of its thirty-first agent, used for a different purpose. There it tests whether a system is any good. Here it tests whether the question can be answered at all — which is the cheaper thing to find out first.

Define a corpus. Don't clean an enterprise.

The instinct after all this is a data quality programme. Resist it. Enterprise-wide cleanup is a multi-year commitment that will outlive the sponsor and the use case.

The alternative is smaller and it works: **for one use case, define one corpus.** Name the documents that are authoritative. Give each an owner and a date it was last verified. Mark everything else out of scope — not deleted, just not indexed. Write down the precedence rule you discovered in the ten-question test, because it is now the most valuable artifact you have.

That is days of work, not quarters. And it's work only the business can do — nobody in IT can rank which of four policy documents wins, because that ranking is domain knowledge, not metadata. It's the same shape as every other real move in enterprise AI: small, ownable, and available without asking anyone's permission.

Then the honest test of whether it worked: run the ten questions again, and see whether the answers now come with a source you'd be willing to defend.

If you want a read on where your organisation stands more broadly, my free AI Readiness Score takes about ten minutes — 20 questions across pilots, data, talent and governance. Two of them are this essay: whether you know the quality and lineage of the data your AI systems consume, and whether your documents, tickets and knowledge bases could feed a use case tomorrow.

Read the essay online at https://burakgaliba.com/blog/documents-are-not-ai-ready

I built a free AI Readiness Score — 20 questions across pilots, data, talent and governance, about ten minutes, no email required: https://burakgaliba.com/readiness