Engineer Data

Company data ready for AI agents

Connect company records to an agreed AI workflow, with clear data sources, quality checks, permissions and a traceable first result.

Give an AI workflow a dependable route to the company records it needs. We map the relevant sources, clarify what each field means, connect the data and set access rules with you. The first deliverable is one agreed question answered from traceable records, with checks for data quality and permissions.

Turn the service into a result

Make “Company data ready for AI agents” serve a business outcome. We can connect the technical work to the users, operations and decisions it needs to support. Your records where AI agents can reach them, one version of each customer, nobody shown more than they should, and every answer traced to its source.

Explore our services Get in touch →

Examine the relevant work

Use the related work on this page to examine our approach. Ask about the tradeoffs and the conditions behind each result. Made a geographical data pipeline 50x faster — 50x faster geographical data pipelines

Made a geographical data pipeline 50x faster Get in touch →

Gathered: one place an agent can reach

We build the pipelines that bring your systems together: ETL jobs that run on a schedule, load only what changed and report what they skipped, into a database whose schema names your business in your words — customers, orders, vessels, contracts, cases. Old PDFs, scans and mail archives are turned into searchable text with their source kept beside them, so they can be found rather than only stored. The agent reaches all of it through one described door: an MCP server or an API, never a copy of the database on somebody’s laptop.

Cleaned: one customer, not five

The same customer can be written five ways in five systems. When an agent reads conflicting versions of one fact, it needs a rule for which source to trust. We decide with you which system is the source of truth for each fact, merge duplicates by rules you can read, and keep the merge running, so a new duplicate is caught on the next load rather than at the next audit. Types and constraints in the database refuse a bad row instead of storing it, every column says what it means, and every migration is versioned and checked, so the schema an agent reads is the schema that is deployed.

Permissioned: each agent sees what its user may

An agent that can read everything is a leak with a friendly voice. The rules belong in the data, not in the prompt: roles in the database (RBAC), read-only by default, row-level rules wherever one customer must never see another's, and personal data kept out of reach unless the task needs it — which is also what GDPR asks. The agent acts with its user's privileges, never an administrator's, and every read is logged with the person it was made for. How we lock the agents themselves is its own page.

An agent is only as right **as the data it reads** — so every answer should say where it came from.

Traceable sources and documented transformations make an AI answer easier to assess.

Every answer traced to its source

An answer drawn from company data is worth what its source is worth, so the source travels with it. Every row keeps where it came from and when: the system, the load, the file and the page. An agent can cite the record behind a figure, and a person can open it. A load that skipped rows says how many and why rather than reporting success, and a figure that turns out wrong leads to the load that brought it in, not to a week of guessing. It is the rule our own reporting is built on: a number is never printed that the run did not measure.

The work behind it

Every claim on this page is a method we have run, and each one has its record. We made a geographical data pipeline 50 times faster by fixing its SQL and its data model, and designed and ran six ETL and ELT pipelines across BigQuery, MS SQL and PostgreSQL. We modelled a whole trade in 348 tables across 34 schemas and kept that schema honest across 1,022 migrations behind a drift gate. We generated 495 pages from a read-only SQLite export, and replaced 963 structured-data blocks with one linked graph, which is what an agent reads first on this site.

How an engagement runs

It starts with a map, made read-only: which systems hold what, who owns each, and which questions an agent should be able to answer. Then we take the one question that matters most and build its path end to end — the sources, the cleaning, the permissions and the door — with checks that stop a load when the data drifts. Then the next question. Everything lives in your infrastructure and your repository, through your review, and the agents that read it can be built for your trade or be the ones you already use. It is bought like the rest of our work, on the terms in what it costs.

If the answers your people need are in more places than anyone can name, tell us where your data lives.

The answer your agent needs **is already in your systems** — it has to be made fit to read.

It starts with a read-only map of which systems hold what, and the one question that matters most to you.

Ask for a data map What it costs →

Scope the work before commissioning it

Bring a representative dataset, the questions it must answer and the current bottleneck. We can agree how to measure quality, throughput and recovery.

See what it costs Get in touch →

Talk through your requirements

Tell us what you need from “Company data ready for AI agents”, including the deadline and the team that will maintain it. We can discuss a practical way forward.

Explore our services

welcome@engineer.company +45 50 10 61 84 Copenhagen, Denmark

Take this page with you

Contact card

Carefully Crafted With Love by D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Copyright © 2025 – present · Engineer ApS · All rights reserved.

Printed from https://data.engineer.company/ai-data/