

A Modular AI Operational Platform for a Multi-Entity Professional Services Firm

How creating a company-wide AI operational layer with several AI agents and governance by design enabled a German professional services group to accelerate request processing by **72%**, cut fee proposal preparation time by **65%**, and free each employee from **6 hours** of repetitive administrative work weekly.

Business challenge

Professional services firms in law, finance, and consulting run on trust, context, and responsiveness. In legal, tax, and accounting, the difference between a good client experience and a missed opportunity often comes down to how quickly the right person understands the request, connects it to the right matter, and moves it forward.

But as such companies grow beyond a single practice, the operating model often becomes harder to control. What works for a compact law firm starts to strain when several business units share clients, brand equity, and expertise, but not the same workflows, systems, or view of client history.

Our client, a **Germany-based professional services group with assurance, tax, legal, and management consulting business units**, had reached that inflection point. Around 100 full-time professionals and part-time consultants across the country supported a shared client base, while customer tracking and project management remained fragmented across business units.

The friction showed up in several places at once:

- **Email as the de facto operating system.** Client requests, project updates, document exchanges, internal coordination, and cross-unit referrals all originated and evolved inside email conversations. What they lacked was structured intake, routing logic, and a quick, unified way to track what came in, who owned it, or what happened next.
- **Manual operations.** Even routine requests required attention from senior lawyers, tax advisors, auditors, and consultants, pulling highly qualified professionals into administrative work. Moreover, with most of the proposal fee preparation done by hand, the client's customers waited for days for what competitors with proper automation delivered in hours.
- **Informal and under-structured workflows across business entities.** Cross-unit coordination happened ad hoc rather than through structured processes. A legal client asking a tax-related question would trigger a phone call or a forwarded email, as there was no controlled workflow.
- **Low visibility into projects and matters.** Each team maintained its own matter records in simplified Excel sheets and DMS templates, leaving the company without a unified view of active matters, ownership, and status.
- **AI adoption lagging behind.** While aware that incorporating artificial intelligence into their daily operations was no longer a nice-to-have, the client still lacked a cohesive plan for implementing it.

With these problems snowballing and no in-house expertise to tackle them, the client turned to Instinctools' [AI Center of Excellence](#).

Solution

Considering that context, Instinctools designed the solution around three principles:

- Sensitive client data must remain under the client's control
- AI outputs must be auditable and explainable enough for professional review
- Every AI-assisted workflow must keep a qualified human in the loop

From the beginning, the project was not treated as a chatbot implementation, where another isolated AI tool sits on top of already fragmented workflows. The client needed a **controlled AI operating layer** that could connect inboxes, documents, matter data, internal systems, and human decision-making into one governed workflow.

For law firms, confidential AI solutions with data privacy and compliance built in aren't optional, as lawyers, auditors, tax advisors, and consultants work with sensitive client data and carry personal responsibility for the advice they provide. **AI could help reduce repetitive operational work, but only if every output remained traceable and reviewable.**

The platform was built to support four core capabilities from the start:

- Email intake and matter management
- Employee support
- Quoting support
- Advisory document analysis

01 Choosing a platform approach over one-off AI tools

The client wanted to move quickly, but they also needed the solution to **scale across several business units**. Building one AI tool at a time would have solved individual problems, but it would also have recreated the same fragmentation the company was trying to overcome.

Instinctools proposed a platform-based approach powered by [GENIE](#), our proprietary agentic AI framework. The idea was to build the foundation once (integrations, context layer, agent orchestration, governance, auditability, and data access) and then reuse it across every subsequent AI module.

What exactly did this approach bring to the table?

- **Centralized AI governance inheritance.** With governance, auditability, and compliance controls embedded at the platform level, every new AI solution inherited them automatically.
- **Data platform with adapters.** Pre-built connectors unified the company's DMS, accounting platform, time-tracking tool, and Microsoft 365 services into a single queryable layer.
- **Cross-agent context synchronization.** It is what makes agentic AI for law firms workable in practice: all AI agents operate within a shared context that connects customers, matters, relationships, and services. Besides eliminating duplicate data entry, this approach enables richer operational intelligence over time.

02 Starting with what hurt most: email intake and matter management

The first of the AI agents for the law firm's operational layer our team developed was the email intake and matter management agent.

Much of the company's work began in Microsoft 365 inboxes: new client requests, follow-ups, internal referrals, document exchanges, administrative questions, and updates on existing matters. Our team connected the platform to the client's mailboxes and built a processing pipeline that **frees lawyers, tax advisors, auditors, and consultants from repetitive triage work**.

When an email comes in, the agent:

- Classifies the intent (new engagement, follow-up, internal referral, or administrative request)
- Identifies whether the sender is an existing client or a new prospect
- Determines which business unit should handle the request
- Checks whether the email relates to an active matter or suggests creating a new matter record when no match is found
- Surfaces deadlines, obligations, and required next steps for professional review

The agent processes emails in **German and English**.

Getting the classification right across four business units with different vocabularies was one of the project's tough puzzles. Tax advisors, lawyers, consultants, and assurance specialists may describe similar work differently or use the same term to mean different things. To address this, our business analyst collaborated with domain experts from each unit to map service categories, routing rules, and matter hierarchies. Where the business units adhere to common logic, the system uses shared classifications. Where they diverge, it maintains unit-specific configurations.

The agent was first tested with a champion group of 10 users. Initial classification accuracy reached around 74%. Through few-shot tuning with real company emails and structured feedback from the early adopters, we **raised the accuracy to 87%**, exceeding the client's initial 85% target.

After validation, the email intake and matter management agent was rolled out across all business units. This gave the client a structured intake mechanism and created the foundation for a **big-picture dashboard showing active matters, task ownership, and engagement status across the entire company**. That way, the AI platform brings multiple practice areas into a single workspace, saving managing partners from calling around or digging through email threads to find out what's happening.

03 Broadening AI usage across the company

After the email intake and matter management agent proved stable in production, we began extending the AI systems across the professional services firm's law, consulting, and finance operations.

Employee assistant chatbot in 2 weeks

Before rolling out AI agents for higher-stakes operations like quoting or advising, we wanted **the company's staff to get comfortable with AI-assisted workflows in a low-risk setting**.

We designed and deployed an **assistant chatbot** that gave employees a controlled environment to interact with AI in everyday scenarios. It operates in German and English, matching the company's working languages, and answers routine internal questions about leave policies, expense procedures, onboarding steps, IT access, and more, drawing from the company's HR documentation and operational FAQs. When it can't answer with sufficient confidence, it routes the query to the right person with context already attached, so nobody has to re-explain the question.

Beyond building trust in AI across the workforce, the **chatbot absorbed the stream of repetitive questions that used to land on managers and office admins, freeing each for up to 6 hours of meaningful work a week**.

- Reading the client request and mapping it to the company's services (tax structuring, legal review, due diligence, etc.)
- Finding the right people for the job based on skills and experience
- Checking who's free through the client's time-tracking system
- Drafting a fee proposal based on the company's rate cards for manager review

What used to be an email back-and-forth across partners, associates, and office admins now runs through a **single structured workflow**, with AI laying the groundwork and the professionals making the final call.

Advisory agent in 4 weeks

Due diligence packages, regulatory filings, and contract portfolios are the kind of high-stakes documents where volume makes manual review unacceptably slow. To ease the daily burden of the client's advisory and assurance teams, we built an agent that took on the heavy lifting. It runs within the AI platform built specifically for law firm operations:

- Ingesting PDF, DOCX, and XLSX documents up to 200 pages long
- Producing summaries at the level of detail the professional needs, from a quick overview to a section-by-section breakdown
- Cross-referencing findings across multiple documents to surface non-obvious patterns, contradictions, and gaps

Every AI-generated finding links back to the exact source document, page, and section, so advisory and assurance staff can **verify the evidence in seconds**.

On the data handling side, the agent **classifies documents by sensitivity level**:

- Publicly available materials like regulatory filings, published financial statements, and industry reports are processed through the cloud inference layer under the existing no-retention agreement.
- Highly sensitive documents, such as materials covered by attorney-client privilege or pre-announcement M&A data, are processed entirely on-premises and never leave the client's servers.

Banking on GENIE showed in the numbers

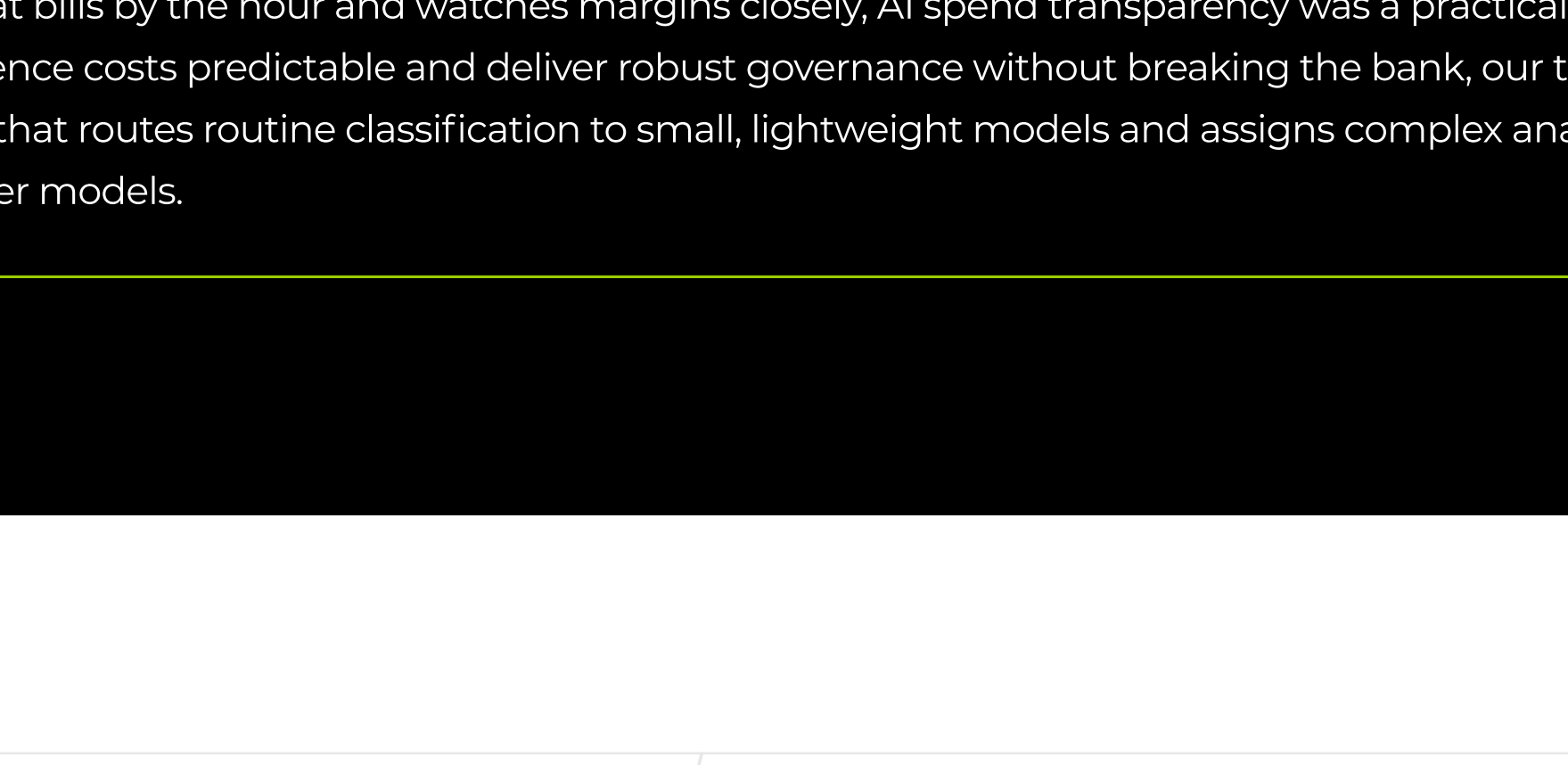
Setting up the governance framework and platform foundation took six weeks. The **multi-agent system** with the assistant chatbot, email intake, quoting, and advisory agents was ready in **14 weeks, bringing the entire delivery timeline down to five months**, well ahead of the client's original 8-10-month estimate.

04 Cost-effective governance by design

For the platform to operate safely across all business units, we set up a robust [AI governance framework](#) from the get-go.

- All application servers, databases, and data pipelines **run on the company's own infrastructure**. Customer data, matter records, and document embeddings never leave the premises.
- Only anonymized, tokenized fragments are sent to Azure OpenAI EU West for inference, under an enterprise agreement that guarantees **no retention or training on the company's data**.
- The framework defines how AI document review works for law firms operating under strict confidentiality: AI outputs are treated as working drafts until reviewed and approved by a qualified professional. This keeps **human oversight built into the workflow** and reflects professional liability requirements under German laws such as BRAO, StBerG, and WPO, as well as the EU AI Act's human-oversight expectations.
- Every action any AI component takes is captured in a **mandatory audit trail with a 12-month retention period**: what decision was suggested, how confident the model was, and whether a professional approved or overrode it. This data can be pulled at any time in a format ready for regulators and auditors.

Here's how NDAs, engagement letters, and service agreements are handled within the AI operational layer.



The client's priority wasn't "governance at any cost"

For a company that bills by the hour and watches margins closely, AI spend transparency was a practical concern from day one. To keep inference costs predictable and deliver robust governance without breaking the bank, our team built a cost-tiered AI gateway that routes routine classification to small, lightweight models and assigns complex analysis and document generation to larger models.

Before

- Email inboxes as the company's informal operating system led to transparency and coordination challenges
- Manual customer intake, matter handling, and quoting wasted senior professionals' billable time
- Client history, matter context, and referral opportunities stayed fragmented across teams and individuals
- No real-time visibility into active matters, ownership, workloads, and engagement status

After

- An AI operational layer with strong governance unlocked transparency and seamless cross-unit coordination
- AI agents took over administrative groundwork, freeing professionals for higher-value expert work
- Shared operational context unified customer and matter knowledge across business units
- Managing partners gained real-time visibility into workloads, matter ownership, active engagements, and cross-unit coordination

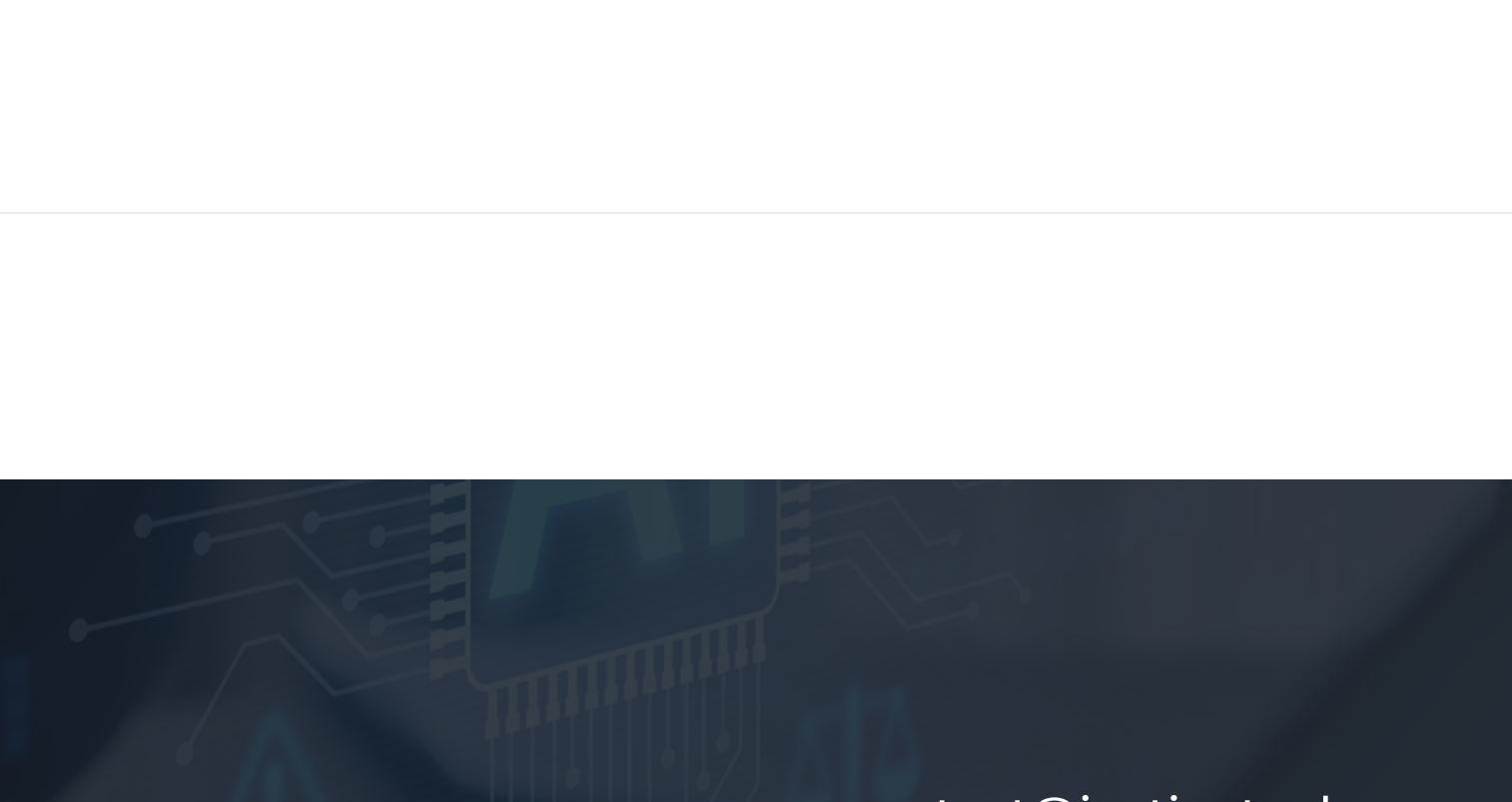
Business value

- **Real-time cross-unit visibility** thanks to a company-wide AI operational layer
- **+72%** in request processing speed
- **-65%** in fee proposal preparation time
- **-6 hours/week** of repetitive administrative work per employee

Multiplier effect

The productivity loss this project addressed goes far beyond legal, tax, and finance. In any knowledge-intensive business, highly qualified people spend too much time on coordination work: routing requests, reconstructing context, chasing updates, preparing drafts, and pulling information from scattered systems.

An AI operational layer with governed multi-agent systems is a way of automating those tasks without sacrificing visibility or control. AI agents take on the groundwork, so your staff spends less time chasing information and more time on the work you hired them for. Companies can start with the biggest time sink and, then, expand gradually as each use case proves its value.



Do you have a similar project idea?

Contact us — and we will estimate your projects costs for free!

[CONTACT US](#)