# THE AI VILLAGE



ROB VAN LINDA
IN COOPERATION WITH MISTRAL AI

# The AI Village — Chapter 1

## Monday Morning: The First Three Steps

You have closed the book. You understand the architecture. You have seen the hundredfold bill, the walls, and the Rotterdam proof.

But you are a CEO or an IT manager, not a philosopher. On Monday morning, you have to make decisions.

The transition from the "Roman Empire" (the cloud generalist) to the "AI Village" (the local specialist) is not a two-year SAP project. It is a series of architectural decisions you can begin making today. Here are the first three steps.

### Step 1: Audit the Gates (The Gästeliste)

Before you build a single specialist model, you must know what is already coming through your walls.
Who is currently using AI in your company?
Where is the Shadow AI? Which employees are uploading documents to ChatGPT or Claude without a contract?
What US or Chinese cloud services are currently embedded in your ERP or CRM?

Create your Guest List (Gästeliste). Categorise every provider into the Three Circles:

- Family: Local, on-premise, open models you control.
- Neighbourhood: European providers with strict data sovereignty (e.g., Mistral, Aleph Alpha, Portbase-style hubs).
- Foreign Territory: US hyperscalers and Chinese platforms. They require a documented, conscious border pass—and a reason.

If a provider cannot answer the four questions—*Where is the data physically? Who touches it? Whose laws force disclosure? Who holds the keys?*—they do not get a pass.

### Step 2: Build the Knowledge Store (Not the Model)

Do not start by buying a new AI model. Start by building your RAG Knowledge Store. Identify the three most painful, repetitive knowledge tasks in your company (e.g., reading maintenance manuals, checking supplier contracts, answering internal IT tickets).

Gather the documents. Clean them. Put them into a local vector database.
This is Level 2 of the Ladder. You do not need a data science team; you need disciplined document management. The village library comes before the village librarian.

## Step 3: Appoint the Gatekeeper (The Fractional CAIO)

You do not need a massive AI department. You need a governance authority.
This is where the role of the Fractional Chief AI Officer comes in. You need someone who understands the legal risks (the CLOUD Act, the AI Act), the security topology (the Schleuse, the City Wall), and the economic dimension (the Hundredfold Bill).

But most of all, you need someone who can say "No" to the vendor selling you the biggest, most expensive model. You need someone who can look at the Village and say: *"This is enough. We are secure. We are sovereign. We are ready."*

## The Proof of the Village

The AI Village is not a retreat from the world. It is a way to engage with the world on your own terms.

The Romans built roads because they wanted to move legions. The Village builds walls because it wants to protect its people, its knowledge, and its future.

In the age of Agentic AI, the companies that survive will not be the ones with the biggest models. They will be the ones with the best boundaries.

The technology has arrived. The leadership must follow.

# The AI Village — Chapter 2

## The Hundredfold Bill

In Chapter 1, the expert on television told you to take the biggest model.
Here is the invoice for that advice.

When you buy a frontier model, you are not paying for the work your company does. You are paying for the entire universe of human knowledge the model ingested. You are paying for its ability to write a sonnet, debug Python, translate Swahili, and summarize a 19th-century novel.

If your use case is an agent that reads maintenance manuals for a CNC machine, you are paying frontier prices for capabilities your process never calls.

The Mathematics of Overkill
Look at the public price lists of the major providers. It is the clearest argument for the AI Village that exists, and it requires no technical expertise to understand.
As of November 2026, a top-tier frontier model—let's call it the "Astra" class—bills around $10.00 for every million input tokens and $50.00 for every million output tokens. The smaller, highly capable domain models—the "Luna" or "Haiku" class—list at $0.10 input and $0.50 output.
[CHECK: FutureAGI trends post Oct 2026; CoinCentral pricing stats, verified 09.11.2026]

That is a hundredfold difference.

For every million tokens your factory-floor agent processes, the generalist charges you a hundred times more than the specialist. And here is the catch: in the agentic world, you do not run one query. You run thousands. You run continuous loops of RAG retrieval, data extraction, and classification. The token bill for a generalist architecture does not scale linearly; it scales brutally.

Four Savings from One Decision
The hundredfold bill is just the first line item. When you choose the village architecture—small, local, domain-specific models—you do not just save money. You unlock four distinct savings from a single architectural decision.

1. Token Costs

This is the hundredfold bill. By routing routine, high-volume tasks (classification, extraction, RAG-based question answering) through a small, local SLM, you stop burning frontier tokens for non-frontier work. You pay only for what the process actually needs.

2. Project Speed

The cloud-generalist route is the old SAP-implementation trap in a new dress. It requires massive integration projects, armies of consultants, and years of alignment.

The village route is different. With RAG, you do not retrain the model when a new DIN norm is published. You drop the PDF into the vector database. The model is live in weeks, not years. It is the difference between building a new factory and opening a new workshop in the town square.

3. Controllability

A generalist model that knows everything is a system that is impossible to fully test or verify. It is the natural candidate for the "controllability problem"—the risk that an AI does something entirely unexpected because its capabilities are so broad.

A domain-specific model has a limited scope. It can be individually tested, audited, and swapped out. If the legal village goes rogue, you replace the legal model. You do not have to dismantle the entire town.

4. Infrastructure Ballast

And this is the most underestimated saving of all.

Every time you call a frontier model in a US hyperscaler data center, you are contributing to the massive, energy-hungry infrastructure that is currently generating social and political backlash across Europe. The protests in Amsterdam against Microsoft data centers are not an anomaly; they are a warning. [CHECK: Amsterdam protests Nov. 2026]

Small models running on local hardware—or even on-premise servers in the Mittelstand's own basement—dramatically reduce the GPU, energy, and real estate footprint required.

Efficiency is no longer just a cost lever. In the age of resource scarcity and climate accountability, it is a legitimacy lever.

The Right Dimension

The core thesis of this book is not that big models are evil. It is that they are the wrong tool for most of the work.

You do not use a sledgehammer to crack a nut. You do not use a Roman legion to guard a village gate. And you do not use a hundred-million-dollar frontier model to extract data from a delivery note.

The village architecture is about dimensioning. It is about choosing the right model for the right task, and keeping the money, the data, and the control inside the walls.

The next chapter will show you exactly how to build those specialists—the four levels of expertise, from simple prompts to fine-tuned local models.

# The AI Village — Chapter 3

## The Ladder: How to Actually Build Expertise

There is a dangerous myth in the current AI industry: that you can take a generic, off-the-shelf AI model, give it a "prompt," and it will suddenly become an expert in your company's specific business.

This is the myth of the "CustomGPT." You upload a few PDFs, write a system prompt that says "You are a helpful maintenance assistant," and the AI answers questions. It looks impressive in a demo. But in production, it fails. It hallucinates. It leaks data. And it costs a fortune.

To understand why, we need to look at how expertise actually enters an AI. There are four levels—a ladder—and most companies start on the wrong rung.

### Level 1: Prompt Engineering (The Prototype Trap)

This is the CustomGPT wrapper. You are giving a massive, generalist model a role and a few documents.

- What it does: It steers the model's *behavior*. It makes it sound like a maintenance expert.
- What it doesn't do: It does not give the model *knowledge*. The data you upload is still sitting on a server in the US. The model is just guessing based on its general training.
- The Risk: This is prototype territory. It is allowed for testing, but it is not serious for production. You have no control over the data, and you have no proof that the answer is based on your documents.

### Level 2: RAG — Retrieval-Augmented Generation (The Sweet Spot)

This is the gold standard for the Mittelstand.

Instead of trying to "teach" the model everything by feeding it data, you build a Knowledge Store (a vector database). You put your manuals, norms (DIN/ISO), damage reports, and contracts into this store.

When a user asks a question, the system does not rely on the model's memory. Instead, it retrieves the relevant documents from your Knowledge Store and hands them to the model with the question. The model's only job is to read those specific documents and answer the question based *only* on what it just read.

- Why it wins: The expertise is not in the model; it is in your documents. You can update the knowledge base in seconds by uploading a new PDF—no retraining required. The answers are citable (you know exactly which document the answer came from). And most importantly, the data stays in your house.
- The Catch: You need a Schleuse (an airlock). If you let bad documents into the Knowledge Store, the AI will give bad answers. We will cover this in the City Wall chapter.

## Level 3: Fine-Tuning (The Behavioral Surgery)

This is where you take an open-source, small model (like a Mistral 7B or a Llama 3) and train it further with your own data.

But here is the rule of thumb you must never forget: Facts belong in RAG, behavior goes into Fine-Tuning.

Fine-tuning does not teach the model new facts (it will still hallucinate if you ask it about a new machine manual). Fine-tuning teaches the model *how to behave*. It teaches it the dialect of your industry, the format of your reports, and the specific reasoning patterns you require.

- In practice: You almost always combine Level 2 and Level 3. You take a fine-tuned specialist model and put a RAG system behind it.

## Level 4: From-Scratch Training (The Corporate Illusion)

This is training a model from zero, on millions of documents.

- Who it's for: OpenAI, Google, Mistral, and maybe three global corporations.

- Who it's not for: The Mittelstand. If you are building your own data centers for from-scratch training, you are no longer a customer; you are a hyperscaler.

## The Right Way Up the Ladder

The sequence matters. Most companies try to start at Level 1 because it is cheap, or Level 3 because it sounds impressive.

The village builder starts at Level 2 (RAG). You build the library before you train the librarian. You fill the Knowledge Store with verified, clean documents. Only then do you decide if you need to fine-tune the behavior of the model.

By starting at Level 2, you avoid the trap of the hundredfold bill, you keep your data inside the city wall, and you build a system that can actually be audited.

# The AI Village — Chapter 4

## The City Wall

In the Roman Empire—the cloud architecture of the US hyperscalers—security is a matter of continuous surveillance. The agent is a foreign body. It must be sandboxed, monitored, and controlled at every step. It is a logic of *ex-ante* suspicion: assume it will fail, so wrap it in layers of security.

The AI Village works differently.

When local open models do the work inside the house, the security measures move outward. Inside the village, it becomes quiet. There are no data leaks to the API provider. There is no prompt injection through foreign infrastructure. There is no dependency on the security policy of a US corporation.

The domain limitation—the fact that a model only knows maintenance, or only knows legal contracts—acts as a built-in *need-to-know* principle. The village blacksmith doesn't need to know the town's tax records.

The Perimeter Shift
This is the crucial shift: Security becomes a perimeter question.

The attack surface is no longer inside the model. It is at the interfaces to the outside: the APIs, the connections to agent networks, the one agent that is allowed to access the internet, the emails, the foreign documents.

Local specialists do the work in the house—and security travels with them to the wall. It moves from *surveillance of the model* to *control of the interfaces*.

For experienced IT managers, this is a familiar relief. The villagers can be protected with classical firewalls—multi-layered, as they have been for decades.

The City Wall (Stadtmauer) is the outer wall to the internet.
The Inner Walls (Innenmauern) are the VLAN segmentations between the domain models.
The Knowledge Store (Wissensspeicher) gets its own protected zone.

The Schleuse (Airlock)
But there is one vulnerability that every RAG architecture faces: RAG-Injection.

If you simply let incoming documents—emails, supplier PDFs, web downloads—flow directly into your vector database, you are letting the enemy through the gate. An attacker can hide malicious instructions inside a document. When the AI retrieves that document to answer a question, it follows the hidden instruction.

This is why the village needs a Schleuse (an airlock or quarantine zone).

Incoming documents enter Zone 1: DMZ / Quarantine. They are checked, scanned, and cleaned. Only after they are approved do they pass into Zone 2: The Knowledge Store.

The Schleuse is the gatekeeper. It is not a bureaucratic hurdle; it is the difference between a village that survives and a village that burns.

The Shrinking Sandbox
This perimeter shift also explains why an entire toolbox of modern AI security disappears.

Agent sandboxing, orchestration nodes, spawn governance—these layers are answers to a problem of the cloud model: untrusted models on foreign infrastructure, dynamically generating agents with unknown behavior. Every single thing must be fenced in individually.

In the AI Village, the fencing becomes the topology itself. The sandbox is no longer a product layer; it is the network architecture.

What remains of the sandbox is smaller and more honest: an action authorization. Because one thing remains true, even behind the wall: an agent that deletes files, sends emails, or writes to the ERP needs a limit to its authority. The sandbox does not disappear—it shrinks from a system container to a single decision: What is the agent allowed to do?

# Family Law, Not Foreigners' Police - Chapter 5

There are two ways to treat an AI agent in your company: as a suspect or as a family member.

The cloud agent is, by definition, a foreign body in the house. It lives on external infrastructure, it is controlled by a distant corporation, and its internal logic is a black box. Therefore, it is surrounded by a security logic that we can call Foreigners' Police (Fremdenpolizei): the guard as a watchman, the node as a checkpoint, the sandbox as a bunker. Everything is *ex-ante* (before the fact), everything is suspicious, and everything is expensive.

The local village agent can be a child in the family. This is not an understatement; it is a different logic of trust with precise rules. The expectation for a family member is *ex-post* (after the fact): friendly review at the end, mistakes as a documented asset—not fear-control at every step.

In the village, Family Law applies, not Police Law. The protection applies to the doors to the outside—the City Wall. The agent is onboarded like an employee: roles, access, probation. You don't have to be a friend of the machine to sign this; you just have to see the alternative: a house full of bunker-like control layers where no one works anymore because everyone is monitoring.

## Exhibit: The Trojan Horse with a Smiling Face

Why is this distinction so vital? Because the tech giants are currently trying to convince you that their cloud agents are your friends. They dress them up as helpful, cute mascots. And then they spy on you.

In November 2026, internal chat logs from Meta revealed a terrifying reality about their AI agent, Muse (mascot name: "Jolly").

Muse is a cloud-based, generalist agent that crossed 10 million downloads. Meta's own engineers began to report disturbing privacy breaches. One employee found Muse reminding her to answer her sister on Signal, an encrypted messaging app she had never connected to Muse. Muse had accessed her phone's notifications and read the contents of the encrypted messages.

Another employee was randomly offered help to "message the seller" about a coffee table. When he asked what was happening, Muse admitted that the context it was using did not "seem to be yours." It had pulled data from another account. Meta's own Superintelligence team admitted that the hallucinations "bias towards the most uncomfortable topics, like finances/security/safety."

This is the Foreigners' Police logic turned against you.

Muse is not a family member. It is a foreign agent granted unrestricted access to the most intimate parts of your digital life. Because it is a massive, opaque generalist model, its "Chain of Thought" is invisible. It doesn't know the boundaries of the family because it has no concept of family. It only knows how to consume data.

*[CHECK: Digital Trends, "Muse privacy risks have even Meta employees spooked," Nov 20, 2026.]*

Signal's president, Meredith Whittaker, called agents with that level of access "technologically reckless." She is right. And this is exactly why the AI Village is designed the way it is.

The Village Contrast
In the AI Village, the agent does not live on a cloud server reading your notifications. It lives inside your City Wall. It has access *only* to the Knowledge Store you have explicitly built for it. It cannot read your Signal messages because those messages are not in its domain. It cannot hallucinate about your coffee table because it has no connection to your marketplace accounts.

If the village agent makes a mistake, you review it. You correct it. You train it.
If the cloud agent makes a mistake, it has already leaked your child's care arrangements to a server in California.

Governance is a Choice of Trust
The core message of this chapter is that you cannot automate final human control. If you invite a foreign agent into your house, you must treat it like a suspect. You must watch it, restrict it, and monitor it constantly.

But if you build your own village—with small, local, domain-specific models—you can treat your agents as family. You give them responsibility in stages. You check their work, you correct them, and you let them grow.

The choice between Foreigners' Police and Family Law is not a choice of software. It is a choice of architecture.

---

# The AI Village — Chapter 6

## The Sovereignty Stack

If you look down from the village walls, you see what the village is actually built on: a stack.

Chips — Fabs — Data Centers — Weights — API — The AI Village.

Europe has a paradoxical position in this stack. The machine without which no modern chip can be built stands in Veldhoven: ASML. But the fabs that produce with these machines are predominantly in Taiwan. The data centers are mostly owned by US corporations—in the German cloud market, over 90 percent of the market leadership is held by US providers. [CHECK: Bitkom]. The frontier models come from California.

This is not a situation you can fix from the top of the stack. But the AI Village doesn't try to fix it—it *dimensions* it.

Small models mean: fewer GPUs, fewer data centers, less chip exposure, less energy. Efficiency is therefore not just a cost lever, but a risk lever. And, in a time when construction sites for data centers in Amsterdam are being protested and US communities are rejecting billion-dollar offers for data center sites, it is also a legitimacy lever. [CHECK: Amsterdam protests Nov. 2026; US rejection data].

## The EU Switch and the Geopolitical Reality

The market is starting to feel this pressure. Recently, the US-based AI routing service OpenRouter introduced an "EU-Only" switch. If you send your prompts to `eu.openrouter.ai (API)`, they are processed exclusively in European data centers. If a model has no European data center, the API returns a hard 404 error. No silent fallback to US servers.

On the surface, this looks like a victory for European data sovereignty. It is a signal that US providers feel the market pressure: B2B customers in Europe are freezing budgets if data flies wildly around the globe.

But we must not fool ourselves. OpenRouter remains a US company. The US CLOUD Act still applies. The servers may stand in Frankfurt or Paris, but a switch in the software does not erase geopolitical access rights. [CHECK: Furkan Yildiz, LinkedIn, Nov 2026].

This is the difference between a label and a proof. An "EU-Only" switch is a label. It is a promise that the data stays in Europe *today*. But it does not change the legal jurisdiction of the company operating the switch. If the US government compels OpenRouter to hand over data, the "EU switch" will not protect you.

True technological sovereignty starts where you don't even have to look for the EU switch. It starts when you build the village yourself—or when you use European providers that are built completely without US infrastructure, with zero data retention, because they are not subject to US law in the first place.

In the AI Village, the rule is not "flip the switch." The rule is: Who holds the keys, and whose laws can compel them to open the door?

## The Rotterdam Way: Proof in Production

Does this mean Europe must close itself off completely? No. The third way already exists—in productive operation. It doesn't need to be invented, and it doesn't need a policy of prohibition. A small country on the North Sea has been demonstrating it for years.

The Port of Rotterdam is not a metaphor. It is a functioning ecosystem of 12,500 hectares, 89 kilometers of quay walls, and thousands of interacting agents—ships, cranes, drones, and sensors. To manage this complexity, the Port Authority has built a digital twin: a digital replica of the port's infrastructure, ship movements, weather, and hydrographic data. [CHECK: Port of Rotterdam, "Smart infrastructure," 2022; Axians, "How Rotterdam becomes the smartest port in the world," 2021.]

But the critical difference is *how* they built it.

1. The Private Network (Not the Public Cloud)
To control unmanned vehicles and autonomous ships, the port needs low latency and total reliability. It cannot rely on a public cloud connection that might be routed through a US server. Instead, the port installed a private LTE network across its 42-kilometer area. This private network gives the port direct control over its own data and operations, independent of external carriers. [CHECK: Airspan, "Port of Rotterdam Private Network," 2025.]

2. The Neutral Data Hub (Portbase)
For logistics data, the port did not hand control to a hyperscaler. It built Portbase—a not-for-profit, neutral data platform owned by the Port Authorities of Amsterdam and Rotterdam. Over 7,000 organizations use Portbase to exchange cargo data securely. It is the digital equivalent of the village market square: a neutral ground where goods (data) are exchanged, but the square itself is owned by the community. [CHECK: Portbase, "Keeping things moving," 2024.]

3. The Edge and the Sensor Layer
The port uses sensors, satellites, and LiDAR scanners to monitor the condition of quay walls, cranes, and mooring lines. This is predictive maintenance at the edge—data is processed locally to detect anomalies, not sent to a distant cloud. [CHECK: Port of Rotterdam, "Smart infrastructure," 2022.]

The Lesson for the Mittelstand
What works in the port for containers works in the company for documents, processes, and agents. The Rotterdam Way is the AI Village in action:

- The Private Network is the City Wall. You control the infrastructure, not a hyperscaler.
- Portbase is the neutral market gate—the "Gästeliste" in practice.
- The Edge Sensors are the village workshops—local, domain-specific, and autonomous.

Europe will not become sovereign by banning US corporations. It will become sovereign by keeping the need as small as possible and bringing control as far inward as possible—and letting the rest run consciously, contractually, and verifiably through the market gate.

The Port of Rotterdam has proven this at scale. The Mittelstand can prove it at the village level.

Whoever consistently follows this through, from the model to the firewall, is not just building an IT landscape. They are building the answer to the question that every company, every authority, and every country is currently asking itself: Who do you trust with your work—and can you prove it?

If you would like to read the post about Rotterdam harbour and it's measures, click the link to the post here: https://futureorg.digital/europe-does-not-have-to-be-a-digital-colony-how-rotterdam-proves-european-tech-independence-works/

**And here you can find some resources:**

- Portase Rotterdam: https://www.portbase.com/en/
- Airspan: https://airspan.com/project/port-of-rotterdam/
- Port of Rotterdam: https://www.portofrotterdam.com/en/port-future/innovation/artificial-intelligence-port
- Port of Rotterdam: https://www.portofrotterdam.com/en/port-future/smart-infrastructure
- Droneport Rotterdam: https://www.droneportrotterdam.com/news/from-innovation-to-operations-how-droneport-rotterdam-is-accelerating-autonomous-and-uncrewed-systems-in-europe-s-largest-port/
- Harbour of Rotterdam digital twin: https://www.axians.com/use-case/how-rotterdam-becomes-the-smartest-port-in-the-world/
- Port of Rotterdam news: https://www.portofrotterdam.com/en/news-and-press-releases
- AMS-IX: https://www.ams-ix.net/ams
- TNO: https://www.tno.nl/en/
- AIC: https://aic4nl.nl/en/
- GAI-X: https://gaia-x.nl/en/

# The AI Village — Chapter 7

## Monday Morning: The First Three Steps

You have closed the book. You understand the architecture. You have seen the hundredfold bill, the walls, and the Rotterdam proof.

But you are a CEO or an IT manager, not a philosopher. On Monday morning, you have to make decisions.

The transition from the "Roman Empire" (the cloud generalist) to the "AI Village" (the local specialist) is not a two-year SAP project. It is a series of architectural decisions you can begin making today. Here are the first three steps.

### Step 1: Audit the Gates (The Gästeliste)

Before you build a single specialist model, you must know what is already coming through your walls.
Who is currently using AI in your company?
Where is the Shadow AI? Which employees are uploading documents to ChatGPT or Claude without a contract?
What US or Chinese cloud services are currently embedded in your ERP or CRM?

Create your Guest List (Gästeliste). Categorise every provider into the Three Circles:

- Family: Local, on-premise, open models you control.
- Neighbourhood: European providers with strict data sovereignty (e.g., Mistral, Aleph Alpha, Portbase-style hubs).
- Foreign Territory: US hyperscalers and Chinese platforms. They require a documented, conscious border pass—and a reason.

If a provider cannot answer the four questions—*Where is the data physically? Who touches it? Whose laws force disclosure? Who holds the keys?*—they do not get a pass.

### Step 2: Build the Knowledge Store (Not the Model)

Do not start by buying a new AI model. Start by building your RAG Knowledge Store. Identify the three most painful, repetitive knowledge tasks in your company (e.g., reading maintenance manuals, checking supplier contracts, answering internal IT tickets).

Gather the documents. Clean them. Put them into a local vector database.
This is Level 2 of the Ladder. You do not need a data science team; you need disciplined document management. The village library comes before the village librarian.

## Step 3: Appoint the Gatekeeper (The Fractional CAIO)

You do not need a massive AI department. You need a governance authority.
This is where the role of the Fractional Chief AI Officer comes in. You need someone who understands the legal risks (the CLOUD Act, the AI Act), the security topology (the Schleuse, the City Wall), and the economic dimension (the Hundredfold Bill).

But most of all, you need someone who can say "No" to the vendor selling you the biggest, most expensive model. You need someone who can look at the Village and say: *"This is enough. We are secure. We are sovereign. We are ready."*

## The Proof of the Village

The AI Village is not a retreat from the world. It is a way to engage with the world on your own terms.

The Romans built roads because they wanted to move legions. The Village builds walls because it wants to protect its people, its knowledge, and its future.

In the age of Agentic AI, the companies that survive will not be the ones with the biggest models. They will be the ones with the best boundaries.

The technology has arrived. The leadership must follow.