

1. System Overview

The big picture — what ALAI is and how the pieces fit together

- [1.1 What is ALAI?](#)
- [1.2 How Everything Connects](#)

1.1 What is ALAI?

What is ALAI?

ALAI Holding AS is an **AI-first company factory**. Instead of a traditional software company with developers, ALAI operates a system of AI agents organized into virtual companies that build, review, secure, and deploy software products autonomously.

Core Principle

“ALAI AS is not a holding company that happens to have a build pipeline. ALAI AS **IS** the build pipeline. The pipeline is the product.”

The Machine

The system runs on two Mac Studio M3 Ultra machines:

Machine	Codename	RAM	Role
Mac Studio 1	ANVIL	96 GB	Infrastructure — databases, Docker, daemons, orchestration, Cloudflare tunnels
Mac Studio 2	FORGE	256 GB	Compute — AI inference, heavy models (72B+), agent worker pool

Connected via 10Gbps Thunderbolt bridge (10.0.0.1 ↔ 10.0.0.2).

Key Components

ALAI System

|

- └─ Mission Control (MC) – Task database (SQLite) – the single source of truth
- └─ Pi-Orchestrator – Daemon that pulls tasks from MC, classifies, routes, executes
- └─ Virtual Companies (8 active) – Specialized AI agent teams with souls, configs, blueprints
- └─ Pipeline Engine – BUILD → REVIEW → SECURITY → OPS → DOCS chain
- └─ Minion System – One-shot autonomous agents for task execution
- └─ HiveMind – Shared knowledge base with semantic search (Qdrant)
- └─ Ollama Fleet – Multi-host model serving (qwen, kimi, llama, custom models)
- └─ BookStack – Documentation wiki (this!)

Products Currently Built by the System

Product	Description	Status
Drop	Digital banking / remittance platform (PSD2-compliant)	Active development
Plock	AI-native warehouse management system (Sweden)	Active development
Lobby	AI-native HR platform	Active development
Bilko	Balkan accounting software	Active development
BasicFakta	Invoice/faktura tool	Active development
Tok	Open banking platform	Active development

1.2 How Everything Connects

How Everything Connects

The Flow: From Idea to Done



Model Tier System

Tasks are classified by complexity (1-5) and routed to the appropriate AI model:

Tier	Model	Host	Pipeline	Use Case
1	qwen3:8b	FORGE	ollama-tool	Simple tasks, classification
2	qwen2.5-coder:32b	ANVIL	ollama-tool → ollama-simple	Medium complexity code
3	Kimi K2.5 (240GB)	FORGE	llama-tool → llama-server	Complex tasks, architecture
4	Claude Sonnet	Cloud	claude-cli	Client-facing, critical work
5	Human	—	human-queue	CEO decisions, sensitive ops

Safety Rails

Tasks matching these patterns are **never auto-processed**:

- `CEO / DECISION` — requires Alem's input
- `BankID / Vipps / Finanstilsynet` — regulated services
- `[TENDER]` — procurement (currently paused)