



IPPOCRA

Innovation with Agency.

The Conviction

The most reliable technology is the one that asks nothing of you.

The information that defines us — the documents we produce, the data we generate, the decisions we make — belongs to whoever creates it. Not to platforms. Not to vendors. Not to whoever holds the API key.

This is why we build the other way around. Systems that run where your data lives, that you own fully and transparently. When control stays with you, technology stops being a risk and becomes an asset.

Innovation with Agency. Innovation without agency is just cleverness. Agency without innovation is just caution. We do both, in that order.

Who We Are

Ippocra is an R&D lab and software company. We design and build intelligent systems for enterprises and regulated environments where data sovereignty, auditability, and long-term ownership are non-negotiable.

We are not a consultancy. We do not sell hours. We ship products.

Our work operates on two complementary fronts:

ILAI — Ippocra Local Artificial Intelligence

Local, on-premise AI systems for business operations. Private by architecture, growing with use.

Ippo — Digital Health

Encrypted, family-owned medical records, designed to be shared on the family's terms, not the platform's.

Meet the Founders



Michele Mattioni, PhD

Founder

Leads product vision, business development, and go-to-market strategy. Michele is the technical and product lead behind ILAI — a local, on-premise AI system designed for organisations that cannot send their data to a third-party cloud. He founded Ippocra to make private, sovereign AI accessible. His focus: skill-based AI architecture, local inference, and systems that are auditable by design.



Myrto Kostadima, PhD

Co-Founder

Works as a specialized Forward Deployed Engineer in core competencies, with a focus on Customer Success. Myrto is the technical bridge between Ippocra's architecture and the client's environment: she designs the integration pattern, deploys the system, and ensures the AI operates correctly within the client's infrastructure, data sources, and communication channels. She is the point of contact for POC scoping, technical integration, operational deployment, and ongoing client support.

The Architecture

Ippocra's systems are built on four architectural principles. These are not features. They are constraints that shape every decision we make.

Skill-Based AI

AI capabilities are implemented as discrete, composable skills — not as a single general-purpose LLM with a prompt. Each skill has a defined input, a defined output, and a defined scope. Skills can be tested, audited, and deployed independently. This is the pattern that makes AI integration viable in regulated environments.

Local-First, Data Sovereignty

Systems run where the data lives. On your hardware, in your network. No inference calls to third-party services. No data egress. No telemetry. If your data is sensitive, it stays sensitive — by architecture, not by promise.

Multi-Profile, Persistent Memory

Each user or system context operates within its own profile — its own memory, its own context, its own accumulated knowledge. Profiles are isolated by default. Memory is persistent across sessions. This maps naturally to multi-tenant enterprise deployments.

Agentic Workflows

Skills are not just queried. They are orchestrated. Ippocra's systems execute multi-step workflows: they read, decide, act, and report. This is how AI moves from “chat with your data” to “your data works for you”.

Core Capabilities

Document Intelligence

Pipeline of extraction that transforms real, chaotic documentation into structured, interrogable context. Contracts, reports, technical specifications, operational documents — ingested, parsed, and made queryable without leaving your network.

AI Agent Design & Integration

Design and deployment of AI agents for specific business domains. An agent is a persistent, context-aware system that executes workflows autonomously within a defined scope. We design the skill set, the memory model, the workflow orchestration, and the integration layer.

Persistent Memory Systems

Long-term memory architectures for AI systems that need to accumulate context over time. Memory is structured, queryable, and owned by the client. Not a vector database. Not a cache. A designed, auditable memory model that grows with the system.

Local LLM Orchestration

Routing of sensitive workloads to the smallest capable local model, on-premise. No cloud dependency. No per-token billing. Predictable cost. The model selection layer is abstracted: your system gets the right capability for the task, not the most expensive model available.

Enterprise AI Architecture Consulting

Advisory for organisations evaluating AI integration in regulated or data-sensitive environments. We help define the architectural boundaries: what runs locally, what runs in a private cloud, what never leaves the perimeter. We map capabilities to compliance requirements.

Multi-Tenant AI Deployment

Deployment patterns for AI systems serving multiple users, teams, or organisations within the same infrastructure. Profile isolation, permission boundaries, context separation, and per-tenant memory management.

ILAI

Ippocra Local Artificial Intelligence.

ILAI is our flagship product: a local, on-premise AI system that operates as a persistent, skill-based assistant within your organisation. It runs on your hardware, connects to your existing communication channels (Slack, Teams, email, WhatsApp, Telegram), and executes workflows on your documents and data.

How it works

1 Onboarding

You configure the system with your context: who you are, what your organisation does, where resources and documents live.

3 Growth

Every interaction leaves knowledge on your machine. Over time, ILAI becomes more useful. The system is an asset, not a service.

2 Execution

You assign a specific task. ILAI executes the workflow, produces the result, and reports. You review and correct if needed.

Deployment models

Local

Runs on your own workstation. Full sovereignty. Zero cloud dependency. You own the hardware, the data, and the accumulated knowledge.

Private Cloud

Dedicated infrastructure with certified Zero Data Retention inference. No hardware on your side. Immediate start.

The Research Lab

Ippocra operates as a research lab, not a product factory. We prototype in public, benchmark obsessively, and publish what we learn. Every product is a step along a longer line of experiments. When one matures enough to support real work, it goes to production. The rest stays in the lab, pushing the limits of what local, private intelligence can do.

Research Register

- **Inference** — Orchestration of local models. Routing sensitive workloads to the smallest capable engine, on-premise. Benchmarking, quantization, and model selection for constrained hardware.
- **Documents** — Extraction pipelines that transform real, chaotic documentation into structured, interrogable context.
- **Health Data** — Encrypted family medical records, owned by the family. Designed to be shared on the family's terms, not the platform's.
- **Sovereignty** — Zero-cloud architectures for regulated sectors: no egress, no telemetry, no escape route.

How We Work

1 Discovery

We start with your context: what data is sensitive, what workflows need to be accelerated, what compliance constraints apply. We map the problem space and define the architectural boundaries.

3 Integration

We design the integration pattern: how the AI system connects to your existing infrastructure, communication channels, and data sources. We handle the deployment, the security model, and the operational runbook.

2 Proof of Concept

We build a working system on your hardware, with your data, for a specific workflow. You evaluate the result.

4 Production

The system goes live. We monitor, we iterate, and the system grows with use. You own the hardware, the data, and the accumulated knowledge. We support the operational lifecycle.

Why Ippocra

Private by Design

Sensitive data never leaves your network. No third-party inference, no hidden dependencies, no trust assumptions. If your data is private, it stays private — by architecture, not by promise.

Auditable by Architecture

Every system we ship is readable, auditable, and under your command. Skills are discrete units with defined inputs and outputs. Workflows are traceable. Memory is structured and queryable. This is what we mean by agency.

Local by Default

Software that runs where it needs to, on hardware you control. Predictable costs, zero lock-in, full sovereignty. You can migrate, fork, or change without being held hostage.

Agency, Always

We build for the customer need. Every system we release is under your control. You decide what it does, where it runs, and what it knows.



Ippocra

info@ippocra.com · ippocra.com

Innovation with Agency