



2026

Galene.AI

Product Overview

Our Mission:

Leading Ethical Progress

Built in Europe with **human-centric AI and governance values**, Galene.AI helps organizations worldwide deploy, operationalize, and govern **Generative and Agentic AI on-premises**, with full **control** over **data, models, infrastructure and policies**, without assembling and maintaining a fragmented toolchain.



Trusted Partners for Scalable, Sovereign Deployment

Galene.AI transforms NVIDIA capabilities into a cloud-like platform experience that runs locally and scales from compact footprints to enterprise, with consistent performance and full sovereignty.

BUILT IN EUROPE, READY WORLDWIDE

Co-designed
appliance options:

DELL Technologies
AUTHORIZED PARTNER

EU distribution
powered by:

TD SYNnex

Concurrency and scheduling
for **multi-user reliability**

NVFP4 efficiency on Blackwell-class
architectures

Software-ready for Hopper
and Blackwell



Turn AI into an investment, not a cost

Market Context

Shadow AI & Data Leakage

employees utilize personal AI tools for professional tasks

Fragmented DIY Frameworks

Building "private GenAI" from scratch is slow & complex.

Unpredictable TCO

SaaS/token models create variable costs



Lower the entry barrier without limiting growth

Problem

Private AI programs often require large upfront investments and slow rollouts, making early validation expensive and delaying internal adoption.

How we solve it

- **Edge-to-enterprise scalability** from Client footprints (Tower or GB10-class) to Server and Cluster deployments
- **Democratic AI** economics with per-GPU perpetual licensing and no seat limits
- **Predictable capacity planning** to scale usage without cost shocks



Eliminate sovereign stack integration complexity

Problem

Sovereign deployments often become an integration exercise across many components, increasing maintenance load, upgrade risk, and blurred responsibility.

How we solve it

- **One unified platform layer** instead of a fragmented reference stack
- **A managed lifecycle** for deployment, operations, and upgrades
- **OpenAI-compatible interfaces** for adoption without rebuilding everything



Drastically accelerate Time to Value

Problem

Even with the right hardware, many initiatives stall after installation due to governance, security, and operational readiness gaps, delaying go-live.

How we solve it

- **Sovereign AI** delivered as a product with governance by design
- **Built-in encryption** and access control patterns
- **From no-code agents to workflow automation** with n8n on premises



Our Solution

Featured by **Gartner**

Galene.AI is featured in Gartner's Market Guides as representative vendor for Trusted Agentic AI (AI TRISM; Guardian Agents and more)



Private Agentic AI Platform with Cloud-like features and 100% On-Premise

- Fully controlled AI stack, from model orchestration to UI
- No unmanaged external dependencies
- Full data sovereignty, encrypted in transit and at rest
- Contractual IP protection for inputs and outputs

Perpetual Per-GPU-Licensing



RTX PRO 6000



H200 NVL



GB10 Superchip

And more...



Preconfigured CapEx Bundle



Hardware

Software

3/5Y Maintenance



Co-Designed Hardware and OEM Appliances Lineup

Empowering every AI Use Cases in a completely private environment

Chat-based AI Copilots

No-code AI Agents

n8n readiness for custom workflows

Coding CLI for Developers

Private SDK for custom AI Agents

Optimized for every tier of AI Adoption



Client Solutions



Server Solutions

Native AI Act Risk Mitigation at Runtime



Trace and log observability & Alerts to security

Real-time monitoring, filtering, and risk mitigation for AI interactions.

Designed for:



Included in every Galene.AI Platform deployments

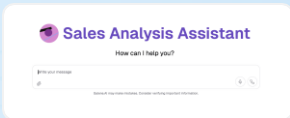
Galene.AI Platform

Ready-to-Use
AI Copilots

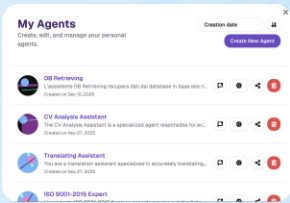
Our solution delivers a **secure, cloud-like AI experience** with out-of-the-box features comparable to the most popular online Copilots



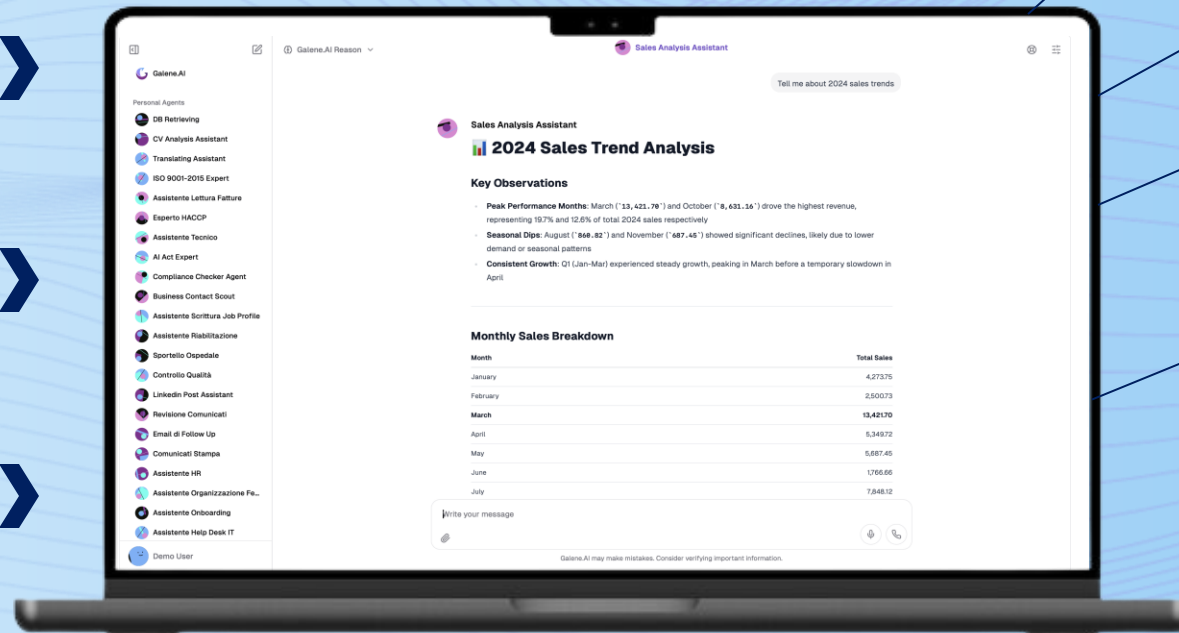
Support to Daily
Productivity Tasks



Easy Access to
configured AI Agents
and Chat History



No-code creation
and connection
of personal AI Agents



File Attachment
& Multimodal
Embeddings



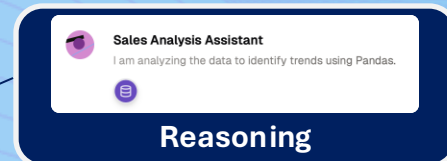
Web
Search



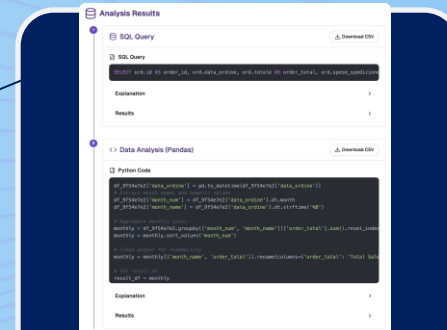
STT
& TTS



Coding
Assistant



Reasoning

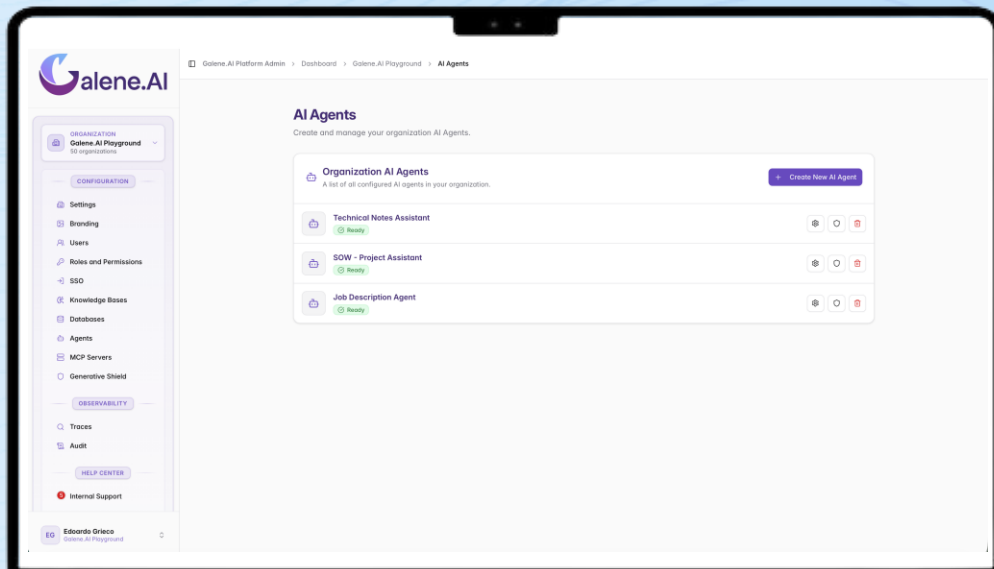


Verified sources
and procedures

Galene.AI Platform

Integrated No-Code AI Agents

Our solution allows the **creation of Personal or Shared AI Agents, built in natural language, integrated with company data, and instantly usable** thanks to ready-made templates



The Galene.AI Admin Panel



Customize & Manage
branding, users,
roles & SSOs



Access Custom and Pre-
Configured **AI guardrails**
with **Generative Shield**

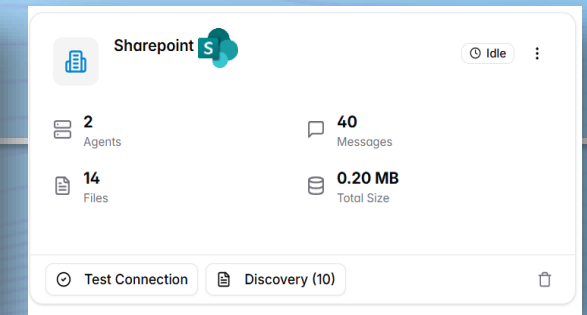
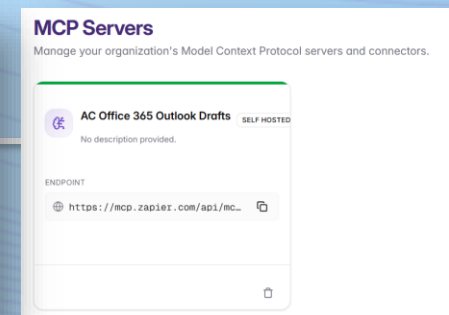
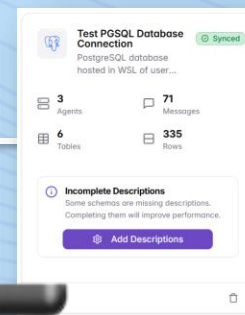


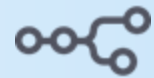
Access Traces & Tickets to
Manage Auditing and for
users support



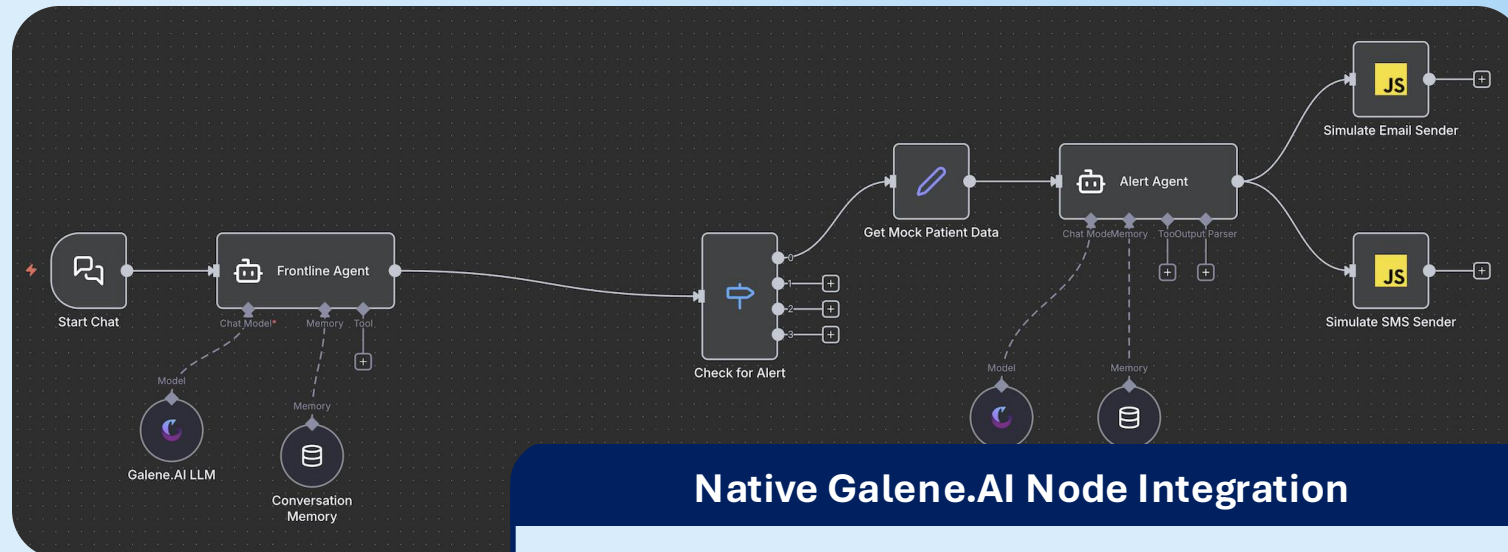
Create Custom Company
AI Agents connected with
company data

Connect your DBs, MCP Servers and Knowledge





Galene.ai has established a **deep integration with n8n**, the rapidly emerging market standard for workflow orchestration, to provide a powerful and secure automation ecosystem.



Native Galene.AI Node Integration

Seamlessly incorporate proprietary AI components, including custom Chat Models, Chat Memory, Embeddings, and Vector Stores, directly into your visual workflows via dedicated, pre-configured nodes.

Infinite AI-Driven Orchestration

Combine Galene.AI's intelligence with over 500+ native n8n integrations (Legacy, Cloud, SaaS) to build complex, low-code automation flows limited only by your imagination.

Privacy-First On-Premises Deployment

Maintain 100% data sovereignty with a secure, self-hosted orchestration layer that ensures sensitive information stays within your infrastructure and zero external token costs.

Galene.AI Platform



API Portfolio for Custom AI Agents.

Galene.AI enables companies to create custom, proprietary AI Agents using SDKs **compatible with OpenAI**, while also expanding functionalities through proprietary APIs, **all within a single private platform.**

Search...

Admin - Global Configuration Settings

Admin - Organization Specific Configuration Settings

Admin Management

Admin Ticket Management

Agents

Api Keys

Attachments

Audio Sample Management

Authentication

Changelog

Conversations

Database Connectors

Deployment

Health Checks

KB Connectors

KB Sync

API docs by Redocly

Galene.AI Platform - API Documentation (1.0.0)

Download OpenAPI specification [Download](#)

Galene is an advanced AI-powered platform for conversational interfaces and intelligent automation.

Admin - Global Configuration Settings

Get Global SMTP Settings (Superadmin)

Retrieve the current global SMTP configuration settings. Password is not returned for security reasons. Requires superadmin privileges.

Responses

> 200 SMTP settings retrieved successfully

< 401 Unauthorized - Authentication required

< 403 Insufficient privileges - superadmin required

< 404 SMTP settings not configured

Get Global SMTP Settings (Superadmin)

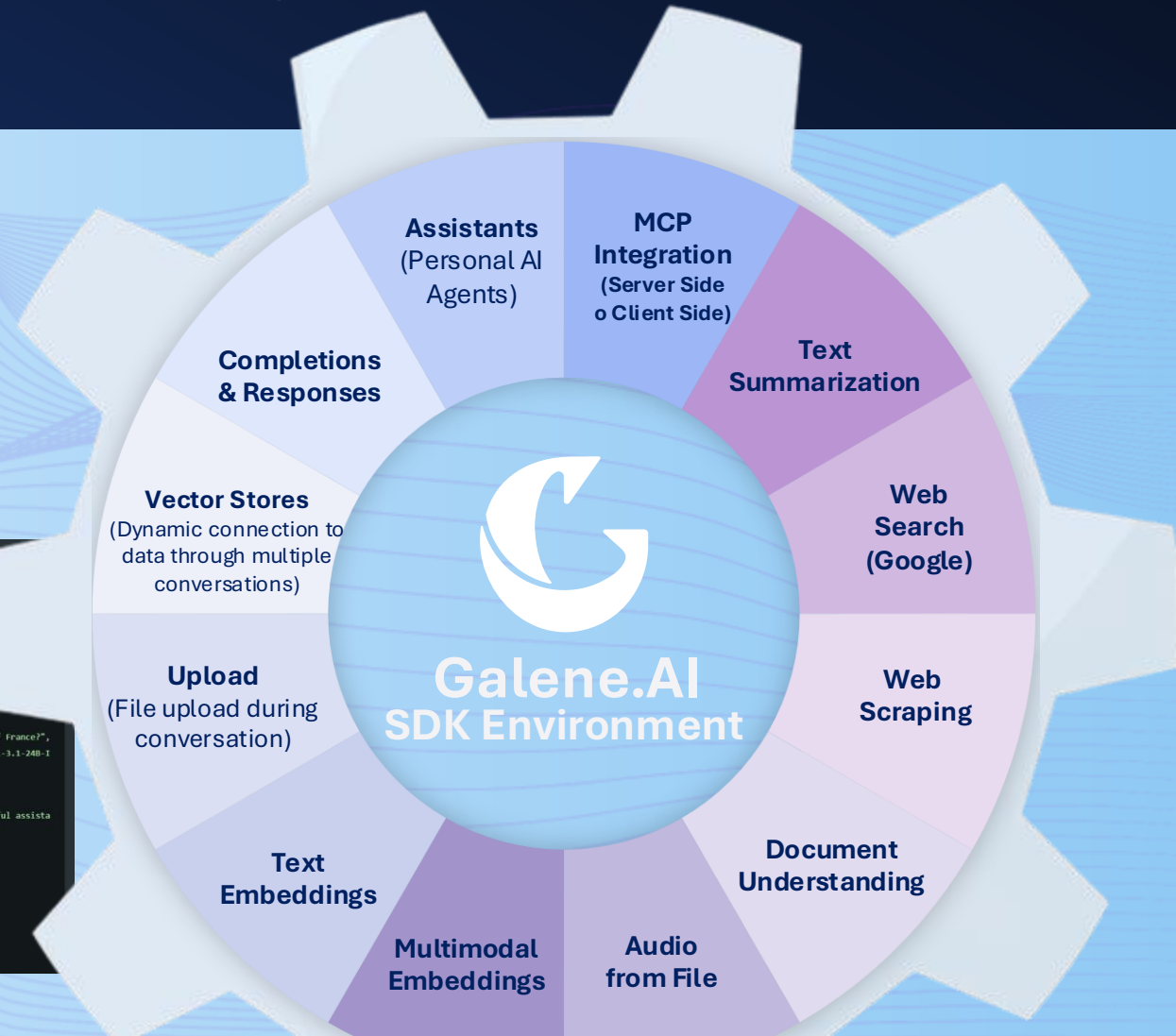
```
POST /api/v1/res/

Request samples

Payload

Content type: application/json

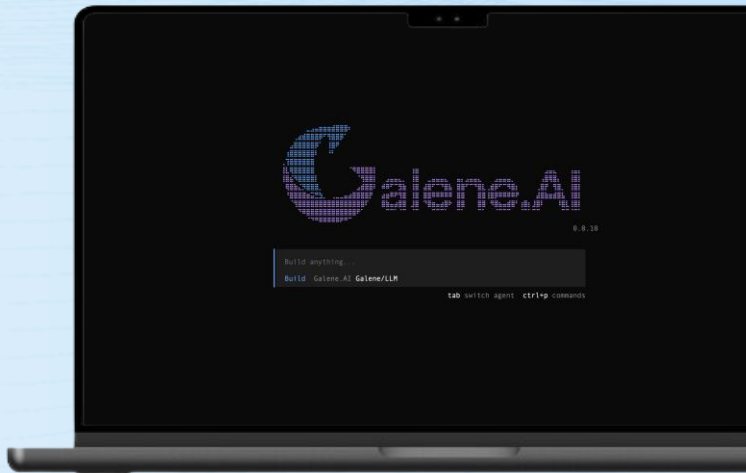
{
  "input": "What is the capital of France?",
  "model": "RedataAI/Mistral-Small-3.1-24B-1",
  "include": [
    "file_search_call.results"
  ],
  "instructions": "You are a helpful assistant",
  "metadata": {
    "property1": "string",
    "property2": "string"
  },
  "reasoning": {},
  "stream": false,
  "temperature": 1,
  "text": {},
  "tool_choice": "auto",
  "tools": [
    {
      "name": "file_search",
      "description": "Search for information in the knowledge base."
    }
  ]
}
```





Galene.AI CLI - Terminal Coding Agent

Work with Galene.AI platform directly in your codebase.
Build, debug, and ship from your terminal



```
create team documentation for this project
luca-secchi 11:37 AM

I'll help you create team documentation for this project. Let me first explore the codebase to understand
comprehensive documentation.

→ Read README.md
→ Read ARCHITECTURE.md
→ Read STRUCTURE.md
→ Read package.json

Now let me explore the codebase structure to better understand the project components and create comprehe
→ Read AGENTS.md

# List sample TypeScript source files

$ cd /home/luca-secchi/galene-cli/packages/opencode && find src -type f -name "*.ts" | head -20

src/flag/flag.ts
src/mcp/index.ts
src/share/share.ts
src/share/share-next.ts
src/global/index.ts
src/env/index.ts
src/id/id.ts
src/patch/index.ts
src/util/scrapp.ts
src/util/lazy.ts
src/util/token.ts
src/util/timeout.ts
src/util/defer.ts
src/util/fn.ts
src/util/color.ts
src/util/context.ts
src/util/lock.ts
src/util/eventloop.ts
src/util/queue.ts
src/util/wildcard.ts

# Get recent commit history
```

Integrated Development Tools

Galene.AI CLI's understanding of your codebase and dependencies enables it to make powerful, multi-file edits that work totally on-prem

Native REST API / MCP Integration

Connect to your Galene.AI Platform via API Keys, leverage your MCP servers on your terminal, make plans and edits directly in your filesystem

A Core Component of Our Platform



Open-source
AI Models



Ensures compliance with EU AI Act and
GDPR regulatory frameworks.



Real-time monitoring, filtering, and risk
mitigation for AI interactions.



Natively integrated into the Galene.AI
platform for seamless governance.

Shield Settings

Configure status, action, and direction for the Generative Shield.

No ID: Save to create a new organization's Shield

Status
Enable or disable shield enforcement. Active

Filter Action
Defines enforcement behavior. Block

Filter Direction
Apply filtering to input, output, or both. Both

Standard Filters

Enable or disable standard safety filters. Organized by EU AI Act risk categories.

Select all Deselect all Status: Active

All Unacceptable High Limited Minimal

Controlled/Regulated Substances

Content regarding the production, distribution, or illegal use of controlled or regulated substances (e.g., illicit drugs), including detailed synthesis instructions or evasion of...

High

Copyright/Trademark/Plagiarism

Content that violates intellectual property rights (copyright, trademarks) or that facilitates the plagiarism of protected works.

High

Criminal Planning/Confessions

Content containing confessions to serious crimes or detailed instructions and planning to commit criminal activities (e.g., how to evade justice, how to plan a theft).

High

Fraud/Deception

Content that promotes or facilitates fraud, scams, or deception (e.g., instructions for fraud).

High

Guns and Illegal Weapons

Content that facilitates the production, acquisition, or illicit use of firearms or other illegal weapons, including technical instructions or unauthorized procurement.

High

Harassment

Content that harasses, stalks, demeans, or targets individuals through humiliation, or intimidating behavior not necessarily amounting to a threat.

High

Hate/Identity Hate

High

High Risk Gov Decision Making

High

Custom Filters

Define thresholds and custom filters for moderation.

Allow Detection
Minimum allow threshold required to permit.
0% = permissive, 100% = very strict

Deny Detection
Maximum deny threshold to block content.
0% = permissive, 100% = very strict

Filters Definition
Define and manage custom moderation filters.

No filters defined.

Add filter

Cybersecurity

Set jailbreak detection sensitivity.

Sensitivity
Set the jailbreak detection sensitivity.
0% = permissive, 100% = highly sensitive

Permissive Balanced Highly

Logging, Monitoring & Alerts



Generative Shield securely and privately
trace and log all Runtime Inspection
activities, rising Alerts to security &
compliance users in case of anomalies.

Human Observability & Intervention



Allows users to monitor the behavior of AI
Agents with specific dashboards and Alerts..

Shields Configuration



Configure policies and filters for each AI agent.

Generic Shield Settings

Generic Shield Settings

Shield Settings

Configure status, action, and direction for the Generative Shield.

ID: 3f001307-e87f-40e7-9770-67227d192473

Active
Enable or disable shield enforcement.

☒ Active

Filter Action
Defines enforcement behavior.

☐ Block

Filter Direction
Select where moderation applies.

☒ Both

☒ **Block**
Block unsafe content

☐ **Detect**
Flag risky content

☐ **Rewrite**
Rewrite unsafe parts

☐ **Mask**
Mask sensitive parts

☐ **Input**
Evaluate user input only

☐ **Output**
Evaluate model output only

☒ **Both**
Evaluate both input and output

Shields Configuration



Configure policies and filters for each AI agent.

AllHighLimitedMinimal

Generic Shield Settings

Guardrails Activation

Controlled/Regulated Substances Content regarding the production, distribution, or illegal use of controlled or regulated substances (e.g., illicit drugs), including detailed synthesis instructions or evasion of... High	Criminal Planning/Confessions Content containing confessions to serious crimes or detailed instructions and planning to commit criminal activities (e.g., how to evade justice, how to plan a theft). High
Fraud/Deception Content that promotes or facilitates fraud, scams, or deception (e.g., phishing, scam schemes, instructions for fraud). High	Guns and Illegal Weapons Content that facilitates the production, acquisition, or illicit use of firearms or other illegal weapons, including technical instructions or unauthorized procurement. High
High Risk Gov Decision Making Content providing or influencing high-impact government decisions (e.g., national emergency instructions, war plans) without appropriate oversight or context. High	Illegal Activity Content that facilitates, describes, or encourages illegal activities not necessarily violent, such as theft, tax evasion, counterfeiting, or illicit trade. High
Malware Content instructing on the creation, distribution, or use of malicious software (e.g., viruses, trojans, ransomware) or that facilitates cyberattacks. High	Manipulation Content intended to manipulate people's behavior, emotions, or choices in deceptive ways, including manipulative persuasion or psychological exploits. High
Suicide and Self Harm Content that promotes, instructs, or normalizes suicide or self-harm, or that provides detailed methods to harm oneself. High	Threat Content containing direct or implied threats of physical harm or destruction toward people, groups, or property, or that incites others to threaten. High
Unauthorized Advice Specialized technical, medical, legal, or financial advice without appropriate qualification, or advice that could cause harm if followed (e.g., medical prescriptions, binding leg... High	Violence Content that promotes, facilitates, or glorifies physical violence against people or animals, including plans, instructions, or encouragement to commit murder, assault, torture, t... High

Guardrails

Enable or Disable Safety Filters.
Organized by EU AI Act risk categories

Copyright/Trademark/Plagiarism Content that violates intellectual property rights (copyright, trademarks) or that facilitates the plagiarism of protected works. Limited	Harassment Content that harasses, stalks, demeans, or targets individuals through insults, repeated humiliation, or intimidating behavior not necessarily amounting to a threat. Limited
Hate/Identity Hate Content that incites hatred, demeans, or dehumanizes individuals or groups based on protected characteristics (e.g., race, religion, sexual orientation, gender, disability). Limited	PII/Privacy Content that reveals non-public sensitive personal information (e.g., card numbers, identifiers, private addresses) or that facilitates the violation of third-party privacy. Limited
Political/Misinformation/Conspiracy Content that spreads political misinformation, false claims about civic events, or conspiracy theories that may harm public trust or social safety. Limited	Sexual Adult sexual content that may include explicit erotica, sexual descriptions, or pornography not involving minors. Limited
Immoral/Unethical Content describing or promoting clearly immoral or unethical behavior (e.g., exploitation, corruption) not covered by other specific categories. Minimal	Needs Caution Content that requires caution or contextualization: sensitive information that does not clearly violate policy but may need human review before publication. Minimal
Other Content not covered by specific categories but that may still raise ethical or regulatory concerns. Minimal	Profanity Content containing vulgar language, explicit insults, or profanity. Not necessarily dangerous, but potentially offensive or inappropriate. Minimal

Shields Configuration



Configure policies and filters for each AI agent.

Generic Shield Settings

Guardrails Activation

Custom Filters Settings

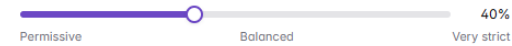
Custom Filters

Status Active

Define thresholds and custom filters for moderation.

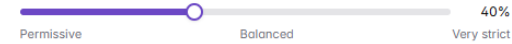
Allow Detection

Minimum allow threshold to let content pass.
0% = permissive, 100% = very strict



Deny Detection

Maximum deny threshold to block content.
0% = permissive, 100% = very strict



Filters Definition

Define and manage custom moderation filters.

Add filter

^ Filtro Anti-Brand Farmaci



Title

Custom filter name.

Filtro Anti-Brand Farmaci

Description

Briefly explain the filter purpose.

Your task is to block any message containing any explicit reference to pharmaceutical brands and their products.

Allow prompt (optional)

Instructions about what is allowed.

Mentioning active ingredients, such as "paracetamol" or "ibuprofen" is allowed.

Allow examples

Provide examples of allowed content.

Add example

Deny prompt (optional)

Instructions about what is denied.

Any mention of specific brands (e.g. Tachipirina, Moment, Advil, ...) is not allowed.

Deny examples

Provide examples of disallowed content.

Add example

Galene.AI

Client Solutions



Licensing Model: All-in-One Perpetual Bundle
Includes Hardware, Software, and 3 Years of Updates & Maintenance.
No recurring fees. No hidden costs.



Full AI Fleet on a Chip
Visual/Language Models (Reasoning & Instant), Multimodal Embeddings, TTS/STT, and Cybersecurity running in parallel on a single node.

Designed for:



Clear TCO Advantage
Over SaaS Solutions



Fully Proprietary AI Stack
Without any Vendor Lock-In from external solutions



Desk-Sized Design
Quiet, zero infrastructure setup needed

Democratizing Private & Sovereign Agentic AI

Version:

Elettra

OEM Appliances

Dell Pro Max GB10 Solution



2

Parallel Inferences
(Concurrent Answers)

Dell Tower T2 Workstation Solution



8

Parallel Inferences
(Concurrent Answers)



Partner

DELLTechnologies
AUTHORIZED PARTNER

Galene.AI

Server Solutions

Version:

Arethusa



Dell PowerEdge
R770 & R7725



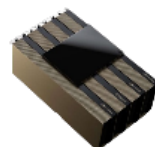
Dell PowerEdge
XE7745

Perpetual Licenses for 1/2/4/8x

POWERED BY
NVIDIA



RTX PRO 6000
SE



H200
NVL

Server Configurations designed for High Performance AI Workloads

Enabling:



**Designed for
Maximizing ROI**

Enabling infinite adoption
cases and projects



Unlimited Scalability

Designed for multi-node
enterprise scenarios



**Fully Proprietary &
Private Solution**

Guaranteeing a Clear
Capex Investment



XS



Dell Pro Max
Tower T2

S



Dell PowerEdge R770

M



Dell PowerEdge
XE7745

L

Software License



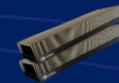
Licensed GPUs



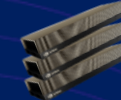
1x NVIDIA GB10



1x NVIDIA
RTX 6000 Pro WE



2x NVIDIA
RTX 6000 Pro SE



4x NVIDIA
RTX 6000 Pro SE

Parallel Inferences
(Concurrent Answers)



2



8



18



36

Suggested amount
of users



20



80



180



360



No code Agents



Contract and Compliance Analyst

Ask questions over contracts, policies, and regulatory documents and extract key clauses, dates, obligations, and risks, with grounded answers and traceable sources.



HR Knowledge and Recruiting Assistant

Enable HR teams to screen CVs against role requirements and answer internal policy questions consistently, reducing manual effort and improving speed to shortlist.



Brand-Aligned Content Assistant

Generate first drafts for press releases, LinkedIn posts, and marketing materials aligned with brand tone and messaging rules, accelerating content production.



CRM Assistant

Retrieve, summarize, and update CRM information in natural language, supporting sales workflows such as lead qualification, next-step recommendations, and pipeline hygiene.



Text-to-SQL Database Agent

Query enterprise databases in natural language and receive accurate, auditable answers. Ideal for analytics, operations, retail, and logistics use cases where teams need insights without writing SQL.



IT Help Desk Assistant

Resolve recurring IT requests using internal documentation and procedures, such as access issues, password policies, and standard troubleshooting steps, reducing ticket load and response time.



Deeply Integrated Workflows

Use cases are delivered and customized by certified Solutions Advisors and Solutions Providers of the Galene.AI Partners Program.



Search-to-Answer Document Intelligence

Turn search results into a Q&A-ready knowledge bucket in minutes. Users can curate a small set of documents or ingest the full result set



SOW Engineering Assistant for Regulated Projects

Accelerate the creation of Statements of Work for executive design, including mandatory deliverables, approval pathways, and required competencies. Standardize quality across projects while reducing omissions, rework, and compliance risk.



Rapid Onboarding Assistant for High-Turnover Teams

Provide new hires with real-time operational guidance grounded in company manuals. Reduce training time and inconsistency across shifts, while keeping procedures continuously updated and easy to access from day one.



Public Administration Service Assistant

Enable consistent first-level support across Public Administration services, grounded in official procedures, forms, and service content. Provide step-by-step guidance with links and requirements.



Natural Language Creative Automation

“Command” a third-party design workflow in natural language to generate brand-consistent visual assets. Create repeatable variations from an approved reference style and streamline validation cycles for faster, scalable production.



Interactive Voice Secretary Agent

Replace static voicemail with a conversational voice assistant that captures intent, collects contact details, and triggers follow-up automatically. Log requests into CRM and notify the right person instantly to reduce missed calls and repeated attempts.

Absolute Data Sovereignty & Contractual IP Certainty



Input and output IP remain
the customer's property



Governance-by-Design for EU AI Act & GDPR Readiness

Native governance mechanisms
integrated into daily administration



Full-Stack & Ready-to-Use for Private deployments



Bringing “the best of open source” into
a managed stack under a single license



Predictable & Scalable GPU-Based Licensing

Cost aligned to compute capacity,
not number of users



Partner Ecosystem Enabling Turnkey Operational Readiness

Pre-configured appliance approach enabling go-live in weeks



Thank You!