Technical Information Sheet
Gaia (IIA France)

Gaia is a platform based on large language models (LLMs - Generative AI) intended for IIA FRANCE members and partners.
Gaia provides specialized conversational agents covering several needs relating to the professional practice of internal auditing. Its main purpose is to provide fast, reliable and secure access to IIA reference materials, assist with the creation of risk matrices, and support the drafting of audit observations and communication materials.

In accordance with Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (the "AI Act"), Gaia falls within the category of limited-risk AI systems. Gaia does not fall within the high-risk system categories listed in Annex III and does not carry out any activities prohibited under Article 5.

Under Article 50 of the Regulation, these systems must comply with transparency obligations toward users when they interact directly with AI, in particular by clearly informing them that they are interacting with an AI system.

As deployer, IIA FRANCE keeps this documentation up to date and makes it available to the authorities upon request[1].

Making this documentation available to users also helps ensure that the system is accompanied by clear and appropriate information.

This technical information sheet covers the following agents:

- Gaia
- Gaia Lex
- Gaia Observation
- Gaia RCM
- Gaia Writer

The Gaia RCM Expert agent is covered by a separate technical information sheet.

[1] In France, the authority responsible for supervising the transparency obligations laid down in Article 50 is the Directorate General for Competition Policy, Consumer Affairs and Fraud Control (DGCCRF).

Contents

1 Definitions

Artificial intelligence system (AI system): within the meaning of Article 3, point 1 of the AI Act, an AI system is "a system based on AI techniques capable, for a given set of human-defined objectives, of generating outputs such as predictions, recommendations or decisions influencing the environments with which they interact".

Deployer: under Article 3, point 4, any natural or legal person using an AI system under its authority, except where the system is used in the course of a purely personal, non-professional activity.

Provider: the entity that develops or places the AI model or system on the market.

General-purpose AI model (GPAI): within the meaning of Article 3, point 63, an AI model that can be used in a variety of applications for general purposes (e.g. GPT-4o). The obligations specific to GPAI models are set out in Article 53 of the Regulation and apply to the provider.

RAG (Retrieval-Augmented Generation): an approach that combines a language model with a structured document repository. Before generating a response, the system searches an internal corpus for the most relevant documents and then provides these materials to the model so that it can produce a contextualized response. This notably improves the reliability and traceability of responses (access to sources).

2 General identification of the Gaia AI system

System name	Gaia (and its agents Gaia, Gaia Lex, Gaia RCM, Gaia Writer and Gaia Observation)
AI model providers	**Mistral:** mistral-medium-3-5 and mistral-embed models. Inference performed on *Mistral AI Infrastructure (Europe)* **Azure OpenAI:** OpenAI models (gpt-4o, gpt-5.4, o3-mini, embeddings-3-large). Inference performed on Microsoft Azure *infrastructure (Europe)*
Deployer	IIA FRANCE
Designated contact	Jean Loup Grosse – Head of Systems and Organization, IIA FRANCE
Hosting	Gaia infrastructure (servers, data): **OVH - France**
Go-live date	May 2025
Current version	Gaia 2.4.1 (May 2026)

3 Description of the Gaia agents covered

Gaia: RAG conversational agent based in particular on the IPPF (IIA International Professional Practices Framework) documents, providing access to sources and referenced responses.

Gaia Lex: conversational agent specializing in the International Internal Audit Standards. Its objective is to provide precise and comprehensive responses based on the 2024 professional standards, with access to sources.

Gaia RCM: agent for the guided development of risk, control and audit test matrices

Gaia Observation: agent assisting with the validation and clear, compliant drafting of audit observations.

Gaia Writer: conversational agent for creating complete materials (articles, communication materials, etc.) using a reasoning model and advanced RAG.

4 Impact, risks and mitigation

Users concerned: IIA FRANCE members and partners with access to the Gaia agents.

4.1 Expected positive impacts

The deployment of the Gaia agents is intended to generate several positive impacts, including:

- Easier access to professional frameworks and standards: users can query the reference corpora directly (IPPF, 2024 Standards, IIA FRANCE documents) through a conversational interface, thereby reducing research and consultation time.
- Development of user skills: through structured, sourced and contextualized responses, Gaia supports users in understanding and adopting internal audit concepts and professional good practices.
- Improved productivity: the agents can assist with certain tasks (initial drafting of documents, identification of standard controls, structured drafting of observations, etc.), freeing up time for analysis and human validation.
- Support for quality and compliance: guided generation, checklists and source-access features reinforce methodological rigor and reduce omissions or inaccurate wording.

4.2 Risks/Potential negative impacts

Despite these benefits, several risks must be taken into account:

- Risk of factual or interpretive errors: some responses may be incomplete, approximate or taken out of context, particularly where wording is ambiguous or because of limitations inherent in AI models.
- Risk of reliance without human verification: users may be tempted to regard responses as accurate without validation, which could lead to erroneous decisions or the dissemination of incorrect information.
- Risk of poor wording or omission of essential elements: particularly when generating audit documents (observations, matrices, materials), certain key components may be missing or poorly structured.
- Risk of overreliance: intensive use of the agents could reduce critical thinking or direct consultation of the reference materials.
- Residual risk of bias or hallucinations: as with any conversational model, the production of incorrect information remains possible, even with an effective RAG system.
- Risk of inappropriate use: the agents could be used for purposes inconsistent with their intended use (for example, outside the intended professional or regulatory framework), or to produce inappropriate or unauthorized content.

4.3 Mitigation measures

To limit these risks, several technical, organizational and methodological measures have been implemented:

- Use of RAG: responses are supported by a validated document corpus (IPPF, standards, IIA FRANCE documents), limiting hallucinations and improving reliability (Gaia, Gaia LEX and Gaia Writer agents)
- Disclaimers: a disclaimer appears at the start of each conversation and is added to every Word/Excel export.
- Instructions: the system prompts for the Gaia agents are designed to limit risks from the outset. They govern the agents' behavior by specializing them in audit and internal control topics, defining the expected types of responses, and incorporating professional good practices directly into their core instructions.
- Guided output: for certain tasks (e.g. RCM matrices), generation follows a structured process and requires final validation before export.
- User training and awareness: prompt-writing assistance is integrated directly into Gaia. IIA FRANCE also offers Gaia training sessions.
- Human review and oversight: the reporting features *(Like/Dislike)* allow problematic responses to be flagged for analysis and continuous improvement.

5 Confidentiality and data protection

5.1 No model training

For each question, the user can choose either a Mistral model or an OpenAI model to generate the response (inference).
The Mistral models use the Mistral APIs on Mistral infrastructure (Mistral AI Infrastructure) in Europe.
The OpenAI models use the Azure OpenAI APIs: OpenAI models on Microsoft Azure infrastructure in Europe.

Prompts (questions) and completions (responses) are not used to improve or train the Mistral[2], OpenAI[3] or Microsoft[4] models.
Likewise, prompts and completions are not used to enrich the IIA FRANCE document repository or the database used by the RAG system.

5.2 Sovereign hosting

The Gaia infrastructure (servers and document repository) is hosted in France in an OVH environment dedicated to IIA FRANCE.

5.3 Data processing and storage

Prompts (questions entered by the user) and responses generated by Gaia are not stored on the Gaia infrastructure[5].

The current conversation for each agent is stored only in the user's web browser and is transferred in full with each new question for stateless processing *(*the conversation is not stored in Gaia).

The user can clear the conversation from the browser's local storage at any time (using the *"New conversation"* and *"Delete all conversations saved locally"* buttons*).*

5.4 Personal and confidential data

No personal data (name, email address, etc.) is stored in Gaia.

[2] https://help.mistral.ai/en/articles/347617-do-you-use-my-user-data-to-train-your-artificial-intelligence-models
[3] https://learn.microsoft.com/en-us/azure/ai-foundry/responsible-ai/openai/data-privacy
[4] https://learn.microsoft.com/en-us/azure/ai-foundry/responsible-ai/openai/data-privacy
[5] With the potential exception of inappropriate content; see "Detection and prevention of inappropriate content".

Only a pseudonymized identifier, generated from the user's login information using a hashing function combined with encryption, is associated with each request. It is used solely for aggregated usage statistics (typically the number of unique platform users).

In accordance with Article 4 of the GDPR, this identifier constitutes pseudonymized personal data. As it stands, it does not allow third parties to identify an individual directly without additional information.
In accordance with the preceding section, "Data processing and storage," any personal or confidential data submitted by the user during use (through prompts) is not stored in Gaia[6].

5.5 Detection and prevention of inappropriate content

OpenAI models (Azure OpenAI – Europe):
A system for analyzing requests (prompts) and responses uses classification models to detect inappropriate content and prevent its generation. These models cover four categories defined by Azure: hate, sexual content, violence and self-harm.

If the system is triggered, response generation is blocked and Gaia displays an error. In the event of repeated detections of abuse, IIA FRANCE may suspend or prohibit access to Gaia for the identifier concerned.
In addition, where the classification model detects inappropriate content, Microsoft may conduct an additional review:

- First, automatically using a more advanced model (LLM)
- Then, where applicable, by a human operator if the model determines that the automated review is insufficient.

The automated review complies with the principles set out in Sections 5.1 (no model training), 5.3 (no storage of prompt/response data) and 5.4 (personal data).
If the case is escalated for human review, the prompt and response are stored by Microsoft for a maximum of 30 days in the Europe region. Microsoft personnel authorized to analyze inappropriate content are located in the European Economic Area.

Except for potentially inappropriate content, prompts and responses are not stored by OpenAI/Microsoft (inference via the Response API without State Storage)[7]

Mistral models:
ZDR (Zero Data Retention) mode is enabled on IIA FRANCE's Mistral API account, and the abuse-detection mechanism is not currently implemented. Consequently, prompts and responses are not stored by Mistral[8].

[6] With the potential exception of inappropriate content; see "Detection and prevention of inappropriate content".
[7] https://learn.microsoft.com/en-us/azure/ai-foundry/responsible-ai/openai/data-privacy
[8] https://legal.mistral.ai/terms/privacy-policy - zero data retention

5.6 Response reporting

Gaia includes a user response-review feature in the form of *Like and Dislike* buttons placed beneath each generated message.
When a user clicks one of these buttons, they indicate the reason and may leave a free-text comment. This comment is stored in the Gaia infrastructure to enable human review by IIA FRANCE. The corresponding conversation is neither transmitted nor stored.

5.7 Usage metrics and logs

The following anonymized metrics are collected and, where applicable, regularly analyzed by the designated IIA FRANCE representative:

- Number of daily uses of each Gaia agent
- Number of unique users
- Breakdown of Gaia usage by intent
- Number of accesses to source documents
- Number of accesses to IIA FRANCE training courses
- Number of response reports
- Number and types of errors encountered
- Inappropriate-content detections
- RAG system performance

To enable these analyses, the following events are logged and may be associated with the non-reversible pseudonymized identifier:

- Use of an agent and associated intent (closed list of 10 intents: *concept/definition,* content *formatting,* risk *& control design, how-to/procedure,* etc.)
- Viewing a source document
- Redirection to an IIA FRANCE training course
- Use of the *Like / Dislike* features
- Error while generating a response
- List of IIA FRANCE documents used to build the response (RAG)

Prompts and responses are never stored in these usage logs.

6 Obligations relating to GPAI models (Chapter V of Regulation (EU) 2024/1689)

The language and embedding models used by Gaia (gpt-5.4, gpt-4o, o3-mini and embeddings-3-large / mistral-medium-3-5 and mistral-embed) are general-purpose AI

models (GPAI models) within the meaning of Article 3, point 63 of Regulation (EU) 2024/1689.

IIA FRANCE is not a GPAI model provider within the meaning of the Regulation. It acts solely as a deployer, using models provided by:

- OpenAI and distributed through Microsoft Azure.
- Mistral and distributed by Mistral.

In accordance with Articles 52 to 55, the preparation and publication of the "model summary" (GPAI information sheet) are the provider's responsibility. These information sheets will be appended to this document as soon as they are officially published.

In the meantime, the public documentation available for each model is listed below:

Model	Technical documentation	System Card
gpt-4o	https://platform.openai.com/docs/models/gpt-4o	https://openai.com/index/gpt-4o-system-card
gpt-5.4	https://learn.microsoft.com/en-us/azure/foundry/foundry-models/concepts/models-sold-directly-by-azure?view=azureml-api-2#azure-openai-in-microsoft-foundry-models	https://openai.com/index/gpt-5-4-thinking-system-card/
o3-mini	https://openai.com/index/openai-o3-mini	https://cdn.openai.com/o3-mini-system-card-feb10.pdf
embeddings-3-large	https://openai.com/index/new-embedding-models-and-api-updates	Not applicable (embedding model)
mistral-medium-3-5	https://docs.mistral.ai/models/model-cards/mistral-medium-3-5-26-04	https://docs.mistral.ai/models/model-cards/mistral-medium-3-5-26-04
mistral-embed	https://docs.mistral.ai/capabilities/embeddings/text_embeddings	Not applicable (embedding model)