Technical Fact Sheet
Gaia (IIA France)

Gaia is a platform based on language models (LLM – Generative AI) intended for IIA FRANCE members and partners. Gaia integrates five specialized AI agents to address key internal audit needs. Its primary objectives are to provide fast, reliable, and secure access to IIA standards, assist in risk matrix development, and support the drafting of observations and communication materials.

In accordance with Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13, 2024, laying down harmonized rules on artificial intelligence ("AI Act"), Gaia falls within the category of low-risk AI systems. Gaia does not fall within the categories of high-risk systems listed in Annex III, nor does it engage in any prohibited activities within the meaning of Article 5.

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

As the deployer, IIA FRANCE keeps this documentation up to date and available to the authorities upon request.

The provision of this documentation to users also aims to ensure that the system is accompanied by clear and appropriate information.

Table of Contents

1 Definitions

Artificial intelligence system (AI system): within the meaning of Article 3.1 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 results such as predictions, recommendations or decisions that influence the environments with which they interact".

Deployer: according to Article 3.4, any natural or legal person who uses an AI system under their authority, unless the system is used in the context of a strictly personal and non-professional activity.

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

General-purpose AI model (GPAI): as defined in Article 3.63, an AI model that can be used in a variety of applications for general purposes (e.g. GPT-4o). The specific obligations for GPAIs are detailed in Article 53 of the Regulation and concern the supplier.

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

2 General Identification of the Gaia AI System

System name	Gaia (5 conversational agents)
Model provider	OpenAI (gpt-4o, o3-mini, embeddings-3-large) on Microsoft Azure infrastructure (Europe)
Deployer	IIA FRANCE
Designated manager	Jean Loup Grosse – IIA FRANCE Systems and Organisation Manager
Commissioning date	May 2025
Current version	Gaia 2.2.0 (August 2025) – Beta version

3 Description of the 5 Gaia agents

Gaia: RAG conversational agent based on IPPF (International Professional Practices Framework of the IIA) documents, providing access to sources and responding with references.

Gaia Lex: Conversational agent specialising in International Standards for the Professional Practice of Internal Auditing. Aims to provide accurate and comprehensive responses based on the 2024 professional standards, with access to sources.

Gaia RCM: Guided construction of risk matrices, controls, and audit tests.

Gaia OBS: Agent that assists in the validation and clear, compliant drafting of audit observations.

Gaia Writer: Conversational agent for the creation of comprehensive 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 Gaia agents.

4.1 Expected positive impacts

The deployment of Gaia agents aims to generate several positive impacts, including:

- Easier access to professional standards and guidelines: users can directly query reference bodies (IPPF, 2024 Standards, IIA FRANCE documents) via a conversational interface, reducing search and consultation time.
- Enhanced user proficiency: thanks to structured, sourced and contextualised responses, Gaia supports the understanding and appropriation of internal audit concepts and professional best practices.
- Improved productivity: agents assist with certain tasks (initial drafting of documents, identification of standard controls, structured writing of observations, etc.), freeing up time for human analysis and validation.
- Support for quality and compliance: guided generation, checklists and access to sources help to reinforce methodological rigour and limit omissions or inaccurate wording.

4.2 Potential negative risks/impacts

Despite these benefits, several risks must be considered:

- Risk of factual or interpretative errors: some responses may be incomplete, approximate or out of context, particularly in the case of ambiguous wording or limitations inherent in AI models.

- Risk of uncritical acceptance: users may be tempted to consider 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, supporting documents), certain key components may be missing or poorly structured.
- Risk of excessive dependence: intensive use of agents could reduce critical thinking or direct consultation of reference materials.
- Residual risk of bias or hallucinations: as with any conversational model, the production of erroneous information remains possible, even with a high-performance RAG.
- Risk of inappropriate use: agents could be used for purposes that do not comply with their intended purpose (e.g. outside the intended professional or regulatory framework), or to produce inappropriate or unauthorised content.

4.3 Mitigation measures

To limit these risks, several technical, organisational and methodological measures have been put in place:

- Use of RAG: all responses are based on a validated body of documentation (IPPF, standards, IIA FRANCE documents), limiting hallucinations and enhancing reliability (Gaia, Gaia LEX and Gaia Writer agents)
- Disclaimers: A disclaimer is included at the beginning of the conversation and added to each Word/Excel export. Instructions: Gaia agent system prompts are written in such a way as to limit risks from the design stage onwards. They guide the behaviour of agents by specialising them in audit and internal control topics, defining the types of responses expected and incorporating best professional practices directly into their basic instructions.
- Production guidance: for certain tasks (e.g. RCM matrices), generation follows a structured process and requires final validation before export.
- User training and support: Gaia includes integrated prompt formulation guidance, and IIA FRANCE offers dedicated training sessions.

- Human review and supervision: reporting features (Like/Dislike) allow problematic responses to be flagged for analysis and continuous improvement.

5 Privacy and data protection

5.1 Non-training of models

The models used are hosted in Europe on Azure infrastructure (AzureOpenAI). Prompts (questions) and completions (answers) are not used to improve or train OpenAI or Microsoft models[1].

Similarly, prompts and completions are not used to enrich the IIA FRANCE document database or the database used by the RAG.

5.2 Hosting

Gaia's infrastructure is hosted on a dedicated Azure environment for IIA FRANCE in Europe (Azure West Europe): data, servers, and LLM inference services.

5.3 Data processing and storage

User prompts (questions entered by the user) and responses generated by Gaia are not stored on Gaia's infrastructure[2].

Each agent's conversation is stored locally in the user's browser and is not retained on Gaia's servers. Conversation history is transmitted with each new query but is not stored.

At any time, the user can clear the conversation from the browser's local storage using the "New conversation" and "Delete all saved conversations locally" buttons.

5.4 Personal and confidential data

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

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

In accordance with Article 4 of the GDPR, this identifier constitutes pseudonymized personal data. It does not, in its current form and for third parties, allow direct identification of an individual without additional information.

In line with the previous section, "Data Processing and Storage," any personal or confidential data transmitted by the user during usage (via prompts) is not stored in Gaia[3].

[1] https://learn.microsoft.com/en-us/azure/ai-foundry/responsible-ai/openai/data-privacy
[2] [3] Except for inappropriate content, see "Inappropriate Content Detection and Prevention."

5.5 Inappropriate content detection and prevention

A request and response analysis system detects inappropriate content and prevents its generation using classification models. These models cover 4 categories defined by Azure: hate, sexual content, violence, and self-harm.

When triggered, the response generation is blocked, and Gaia displays an error. In cases of repeated abuse detection, IIA FRANCE may suspend or revoke access to Gaia for the concerned identifier.

Additionally, if inappropriate content is detected by the classification model, Microsoft may conduct further review:

- Initially, through automated processes using a more advanced model (LLM).
- If necessary, by a human operator if the automated review is deemed insufficient by the model.

The automated review adheres to the principles outlined in sections 5.1 (no model training), 5.3 (no storage of prompt/response data), and 5.4 (personal data). In cases of escalation for human review, the prompt and response are stored by Microsoft (for a maximum of 30 days) in the European region. Microsoft personnel authorized to analyze inappropriate content are located within the European Economic Area (EEA).

5.6 Reporting responses

Gaia includes a user feedback feature for responses, in the form of Like and Dislike buttons placed under each generated message.

When a user clicks on one of these buttons, the current conversation is automatically sent in clear text via email to the designated IIA FRANCE representative to enable human review. No conversation data is stored in Gaia's infrastructure when using this feature; only the email sent contains the conversation.

5.7 Usage metrics and logs

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

- Daily usage count per Gaia agent
- Number of unique users
- Usage distribution of Gaia by intent
- Number of accesses to source documents
- Number of accesses to IIA FRANCE training materials
- Number of response flagging incidents
- Number and typology of encountered errors
- Inappropriate content detections
- Performance of the RAG system

To enable these analyses, the following events are logged, potentially linked to a non-reversible pseudonymized identifier:

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

Except in cases of inappropriate content detection, prompts and responses are never stored in these usage logs, and no personal data is retained in clear text.

6 Obligations Related to General-Purpose AI (Chapter V of Regulation (EU) 2024/1689)

The language/embedding models used by Gaia (gpt-4o, o3-mini, and embeddings-3-large) are general-purpose AI models (GPAI) as defined in Article 3.63 of Regulation (EU) 2024/1689.

IIA FRANCE is not a GPAI provider under the regulation; it acts solely as a deployer, using models supplied by OpenAI and distributed through Microsoft Azure.

In accordance with Articles 52 to 55, the production and publication of the "model summary" (GPAI fact sheet) are the responsibility of the provider. These fact sheets will be appended to this document once officially published.

In the meantime, the publicly available documentation for each model is indicated below:

Model	Technical documentation	System Card
gpt-4o	https://platform.openai.com/docs/models/gpt-4o	https://openai.com/index/gpt-4o-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)