Fiche Technique
Gaia (IFACI)

Gaia est une plateforme basée sur des modèles de langage (LLM - IA Générative) destinée aux adhérents et partenaires de l'IFACI.
Gaia propose des agents conversationnels spécialisés couvrant plusieurs besoins en matière de pratiques professionnelles de l'audit interne. Son objectif principal est de fournir un accès rapide, fiable et sécurisé aux référentiels de l'IIA, d'assister à la création de matrices de risques et d'accompagner la rédaction d'observations ou de supports de communication.

Conformément au Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle ("AI Act"), Gaia relève de la catégorie des systèmes d'IA à risque limité. Gaia n'entre pas dans les catégories de systèmes à haut risque listées à l'Annexe III, et n'exerce pas non plus d'activités interdites au sens de l'article 5.

En vertu de l'article 50 du règlement, ces systèmes doivent respecter des obligations de transparence vis-à-vis des utilisateurs lorsqu'ils interagissent directement avec une IA, notamment en informant clairement qu'ils s'adressent à une IA.

En tant que déployeur l'IFACI tient à jour la présente documentation à disposition des autorités en cas de demande[1].

La mise à disposition de cette documentation aux utilisateurs vise également à garantir que le système est accompagné d'une information claire et appropriée.

Cette fiche technique concerne les agents suivants :

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

L'agent Gaia RCM Expert fait l'objet d'une fiche technique spécifique

[1] En France, l'autorité compétente pour la supervision des obligations de transparence prévues à l'article 50 est la Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes (DGCCRF).

Sommaire

1 Définitions

Système d'intelligence artificielle (Système d'IA) : au sens de l'article 3, point 1 de l'AI Act, un système d'IA est « un système basé sur des techniques d'IA capables, pour un ensemble donné d'objectifs définis par l'homme, de générer des résultats tels que des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels ils interagissent ».

Déployeur : selon l'article 3, point 4, toute personne physique ou morale qui utilise un système d'IA sous son autorité, sauf si le système est utilisé dans le cadre d'une activité strictement personnelle et non professionnelle.

Fournisseur : l'entité qui développe ou met sur le marché le modèle ou le système d'IA.

Modèle d'IA à usage général (GPAI) : au sens de l'article 3, point 63, un modèle d'IA qui peut être utilisé dans une pluralité d'applications pour des finalités générales (ex. GPT-4o). Les obligations spécifiques aux GPAI sont détaillées à l'article 53 du règlement, elles concernent le fournisseur.

RAG (Retrieval-Augmented Generation) : approche consistant à combiner un modèle de langage avec une base documentaire structurée. Avant de générer une réponse, le système recherche dans un corpus interne les documents les plus pertinents, puis fournit ces éléments au modèle afin qu'il produise une réponse contextualisée. Cela permet notamment d'améliorer la fiabilité et la traçabilité des réponses (accès aux sources).

2 Identification Générale du système d'IA Gaia

Nom du système	Gaia (et ses agents Gaia, Gaia Lex, Gaia RCM, Gaia Writer et Gaia Observation)
Fournisseurs de modèles IA	**Mistral** : modèles mistral-medium-3-5, mistral-embed. Inférence effectuée sur *Mistral AI Infrastructure (Europe)* **AzureOpenAI** : modèles OpenAI (gpt-4o, gpt-5.4, o3-mini, embeddings-3-large). Inférence effectuée sur infrastructure *Microsoft Azure (Europe)*
Déployeur	IFACI
Responsable désigné	Jean Loup Grosse – Responsable Systèmes et Organisation IFACI
Hébergement	Infrastructure Gaia (serveurs, données) : **OVH - France**
Date de mise en service	Mai 2025
Version actuelle	Gaia 2.4.1 (mai 2026)

3 Description des agents Gaia concernés

Gaia : Agent conversationnel RAG basé notamment sur les documents de l'IPPF (International Professional Practices Framework de l'IIA), permettant d'accéder aux sources et de répondre avec références.

Gaia Lex : Agent conversationnel spécialisé dans les Normes Internationales d'Audit Interne. Objectif de précision et d'exhaustivité dans les réponses, qui sont basées sur les normes professionnelles 2024, avec accès aux sources.

Gaia RCM : Agent de construction guidée de matrices de risques, contrôles et tests d'audit

Gaia Observation : Agent d'aide à la validation et rédaction conforme et claire d'observations d'audit.

Gaia Writer : Agent conversationnel de création de supports complets (articles, supports de communication, …) avec utilisation d'un modèle de raisonnement et d'un RAG avancé.

4 Impact, risques et mitigation

Utilisateurs concernés : adhérents IFACI et partenaires disposant d'un accès aux agents Gaia.

4.1 Impacts positifs attendus

Le déploiement des agents Gaia vise à générer plusieurs impacts positifs, incluant :

- Accès facilité aux référentiels et normes professionnelles : les utilisateurs peuvent interroger directement les corpus de référence (IPPF, Normes 2024, documents IFACI) via une interface conversationnelle, réduisant ainsi le temps de recherche et de consultation.
- Montée en compétence des utilisateurs : grâce à des réponses structurées, sourcées et contextualisées, Gaia soutient la compréhension et l'appropriation des concepts d'audit interne et des bonnes pratiques professionnelles.
- Amélioration de la productivité : les agents permettent d'assister certaines tâches (formulation initiale de documents, identification de contrôles types, rédaction structurée d'observations…), libérant du temps pour l'analyse et la validation humaine.
- Support à la qualité et à la conformité : les fonctionnalités de génération guidée, de checklists et d'accès aux sources permettent de renforcer la rigueur méthodologique et de limiter les oublis ou formulations inexactes.

4.2 Risques/Impacts potentiels négatifs

Malgré ces bénéfices, plusieurs risques doivent être pris en compte :

- Risque d'erreurs factuelles ou d'interprétation : certaines réponses peuvent être incomplètes, approximatives ou hors contexte, notamment en cas de formulations ambiguës ou de limitations inhérentes aux modèles d'IA.
- Risque de prise en compte sans vérification humaine : les utilisateurs peuvent être tentés de considérer les réponses comme exactes sans validation, ce qui pourrait conduire à des décisions erronées ou à la diffusion d'informations incorrectes.
- Risque de mauvaise formulation ou d'omission d'éléments essentiels : en particulier lors de la génération de documents d'audit (observations, matrices, supports), certaines composantes clés peuvent être absentes ou mal structurées.
- Risque de dépendance excessive : l'utilisation intensive des agents pourrait réduire l'esprit critique ou la consultation directe des référentiels.
- Risque résiduel de biais ou hallucinations : comme pour tout modèle conversationnel, la production d'informations erronées reste possible, même avec un RAG performant.
- Risque d'utilisation inappropriée : les agents pourraient être sollicités à des fins non conformes à leur objet (par exemple, en dehors du cadre professionnel ou réglementaire prévu), ou pour produire du contenu inadapté ou non autorisé.

4.3 Mesures de mitigation

Pour limiter ces risques, plusieurs mesures techniques, organisationnelles et méthodologiques sont mises en place :

- Recours au RAG : les réponses sont appuyées sur un corpus documentaire validé (IPPF, normes, documents IFACI), limitant les hallucinations et renforçant la fiabilité (agents Gaia, Gaia LEX et Gaia Writer)
- Disclaimers : Un disclaimer est présent en début de conversation et ajouté à chaque export Word/Excel.
- Instructions : Les prompts systèmes des agents Gaia sont rédigés de manière à limiter les risques dès la conception. Ils encadrent le comportement des agents en les spécialisant sur les thématiques d'audit et de contrôle interne, en définissant les types de réponses attendues et en intégrant les bonnes pratiques professionnelles directement dans leurs instructions de base.
- Guidage de la production : pour certaines tâches (par ex. matrices RCM), la génération suit un processus structuré et impose une validation finale avant export.
- Formation et sensibilisation des utilisateurs : les utilisateurs disposent d'une assistance à la formulation de prompts directement intégrée à Gaia. Des sessions de formation à Gaia sont proposées par l'IFACI.

- Revue humaine et supervision : les fonctionnalités de signalement (*j'aime/je n'aime pas*) permettent une remontée des réponses problématiques pour analyse et amélioration continue.

5 Confidentialité et protection des données

5.1 Non-entrainement des modèles

A chaque question, l'utilisateur à la possibilité de choisir soit un modèle Mistral, soit un modèle OpenAI pour la rédaction de la réponse (inférence).
Les modèles Mistral utilisent les API Mistral sur l'infrastructure Mistral (Mistral AI Infrastructure), en Europe.
Les modèles OpenAI utilisent les API AzureOpenAI : modèles OpenAI sur infrastructure Microsoft Azure, en Europe.

Les prompts (questions) et complétions (réponses) ne sont pas utilisés pour améliorer ou entrainer les modèles Mistral[2], OpenAI[3] ou Microsoft[4].
De même, les prompts et complétions ne sont pas utilisés pour enrichir la base documentaire IFACI ou la base de données utilisée par le RAG.

5.2 Hébergement Souverain

L'infrastructure de Gaia (serveurs, base documentaire) est hébergée sur un environnement OVH propre à l'IFACI en France.

5.3 Traitement et stockage des données

Les prompts (questions saisies par l'utilisateur) et les réponses générées par Gaia ne sont pas stockées sur l'infrastructure de Gaia[5].

La conversation en cours pour chaque agent est stockée uniquement dans le navigateur internet de l'utilisateur, et est transférée en intégralité à chaque nouvelle question pour traitement sans stockage (*Stateless* - pas de stockage de la conversation dans Gaia).

L'utilisateur peut à tout moment effacer la conversation du stockage local du navigateur (boutons « *Nouvelle conversation* » et « *Supprimer toutes les conversations sauvées en local* »)

[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] A l'exception potentielle du contenu inapproprié, voir « Détection et prévention du contenu inapproprié »

5.4 Données personnelles et confidentielles

Aucune donnée personnelle (nom, e-mail, …) n'est stockée dans Gaia.
Seul un identifiant pseudonymisé, généré à partir des informations de connexion de l'utilisateur via une fonction de hachage combinée à un chiffrement, est associé à chaque requête et n'est utilisé qu'à des fins de statistiques d'utilisation agrégées (typiquement nombre d'utilisateurs uniques de la plateforme).

Conformément à l'article 4 du RGPD, cet identifiant constitue une donnée à caractère personnel pseudonymisée. Il ne permet pas, en l'état et pour des tiers, d'identifier directement une personne sans information supplémentaire.
Conformément au point précédent « Traitement et stockage des données » les éventuelles données personnelles ou confidentielles transmises par l'utilisateur lors de l'utilisation (à travers les prompts) ne sont pas stockées dans Gaia[6].

5.5 Détection et prévention du contenu inapproprié

Modèles OpenAI (Azure OpenAI – Europe) :
Un système d'analyse des requêtes (prompts) et des réponses permet de détecter le contenu inapproprié et d'en empêcher la production, à l'aide de modèles de classification. Ces modèles couvrent 4 catégories définies par Azure : haine, sexualité, violence et automutilation.

En cas d'activation, la génération de la réponse est bloquée et Gaia affiche une erreur. En cas de détections répétées d'abus, l'IFACI est susceptible d'appliquer une suspension ou une interdiction d'accès à Gaia pour l'identifiant concerné.
Par ailleurs, en cas de détection de contenu inapproprié par le modèle de classification, Microsoft est susceptible d'effectuer une revue complémentaire :

- Dans un premier temps de manière automatisée avec un modèle plus évolué (LLM)
- Puis, le cas échéant, par un opérateur humain si la revue automatisée est jugée insuffisante par le modèle.

La revue automatisée respecte les principes 5.1 (non-entrainement du modèle), 5.3 (pas de sauvegarde de données prompt/réponse) et 5.4 (données personnelles).
En cas d'escalade pour revue humaine, le prompt et la réponse sont stockés par Microsoft (30 jours maximum), en zone Europe. Le personnel Microsoft autorisé à l'analyse du contenu inapproprié est localisé dans l'Espace économique européen.

[6] A l'exception potentielle du contenu inapproprié, voir « Détection et prévention du contenu inapproprié »

A l'exception du contenu potentiellement inapproprié, les prompts et réponses ne sont pas stockés par OpenAI/Microsoft (Inférence via *Response API* sans *State Storage*)[7]

Modèles Mistral :
Le mode ZDR (*Zero Data Retention*) est activé sur le compte API Mistral de l'IFACI, et le dispositif de détection d'abus n'est pas implémenté pour l'instant. En conséquence les prompts et réponses ne sont pas stockées par Mistral[8].

5.6 Signalement des réponses

Gaia intègre une fonctionnalité de revue des réponses par les utilisateurs, sous la forme de boutons *J'aime* et *Je n'aime pas* placés sous chaque message généré.
Lorsqu'un utilisateur clique sur l'un de ces boutons, il indique la raison, et a la possibilité de laisser un commentaire sous forme de texte libre. Ce commentaire est stocké dans l'infrastructure Gaia, afin de permettre une revue humaine par l'IFACI. La conversation correspondante n'est pas transmise ni stockée.

5.7 Métriques et traces d'utilisation

Les métriques anonymisées suivantes sont collectées et, le cas échéant, analysées régulièrement par le responsable désigné de l'IFACI :

- Nombre d'utilisations quotidiennes de chaque agent Gaia
- Nombre d'utilisateurs uniques
- Répartition de l'utilisation de Gaia par intention
- Nombre d'accès aux documents sources
- Nombre d'accès aux formations IFACI
- Nombre de signalements de réponses
- Nombre et typologie des erreurs rencontrées
- Détections de contenu inapproprié
- Performances du système de RAG

Pour permettre ces analyses, les événements suivants sont enregistrés, éventuellement associés à l'identifiant pseudonymisé non réversible :

- Utilisation d'un agent et intention associée (liste fermée de 10 intentions : *concept/definition*, *content formatting*, *risk & control design*, *how-to/procedure*, etc.)
- Consultation d'un document source
- Redirection vers une formation IFACI
- Utilisation des fonctionnalités *J'aime / Je n'aime pas*

[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

- Erreur lors de la génération d'une réponse
- Liste des documents IFACI utilisés pour construire la réponse (RAG)

Les prompts et les réponses ne sont jamais stockés dans ces traces d'utilisation.

6 Obligations relatives aux GPAI (Chapitre V du Règlement UE 2024/1689)

Les modèles de langage/embedding utilisés par Gaia (gpt-5.4, gpt-4o, o3-mini et embeddings-3-large / mistral-medium-3-5 et mistral-embed) sont des modèles d'IA à usage général (General Purpose AI models – GPAI) au sens de l'article 3, point 63 du Règlement (UE) 2024/1689.

L'IFACI n'est pas fournisseur de GPAI au sens du règlement : elle agit uniquement en tant que déployeur, en utilisant les modèles fournis par :

- OpenAI et distribués via Microsoft Azure.
- Mistral et distribués par Mistral.

Conformément aux articles 52 à 55, la production et la publication du "résumé du modèle" (fiche GPAI) relèvent de la responsabilité du fournisseur. Ces fiches seront annexées au présent document dès leur publication officielle.

En attendant, la documentation publique disponible pour chaque modèle est indiquée ci-dessous :

Modèle	Documentation technique	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	Non applicable (modèle d'embedding)
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	Non applicable (modèle d'embedding)