Fiche Technique
Gaia RCM V2 (IFACI)

Gaia RCM V2 est un agent basé sur des modèles de langage (LLM – IA Générative) destiné aux adhérents et partenaires de l'IFACI, accessible depuis la plateforme Gaia.

L'objectif principal de Gaia RCM V2 est d'assister les utilisateurs dans l'élaboration de matrices de risques, de référentiels de contrôles et de programmes d'audit. Il s'appuie pour cela sur un référentiel de risques, contrôles et tests d'audit constitué par l'IFACI.

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 RCM V2 relève de la catégorie des systèmes d'IA à risque limité. Gaia RCM V2 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 exclusivement l'agent Gaia RCM V2.

Les agents suivants font l'objet d'une fiche technique spécifique :

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

[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 RCM V2

Nom du système	Gaia RCM V2
Fournisseur du modèle	AzureOpenAI : modèles OpenAI (gpt-4o, gpt-5.1, embeddings-3-large) sur infrastructure Microsoft Azure (Europe)
Déployeur	IFACI
Responsable désigné	Jean Loup Grosse – Responsable Systèmes et Organisation IFACI
Date de mise en service	Décembre 2025
Version actuelle	Gaia 2.3.0 (décembre 2025) – Version Beta

3 Description de l'agent Gaia RCM V2

Gaia RCM V2 est un agent IA accessible depuis la plateforme Gaia de l'IFACI.
Il assiste l'utilisateur dans la préparation du programme de travail et de la matrice risques/contrôles/tests d'audit (RCM), en fonction de son contexte d'audit.

- Une première phase conversationnelle guide l'utilisateur dans la définition de son contexte d'audit.
    - À tout moment dans l'échange l'utilisateur peut valider le contexte d'audit proposé (« je valide le contexte d'audit »).
    - L'utilisateur est alors invité à lancer la génération de la RCM (bouton « Générer la RCM »)
- Un modèle de raisonnement analyse alors ce contexte d'audit et interroge selon différentes stratégies une base de risques/contrôles/tests de référence préalablement constituée. Il peut également proposer des compléments issus de ses propres capacités.
- A l'issue de cette phase (5 à 10 minutes), sont proposés à la revue de l'utilisateur :
    - Une liste d'objectifs d'audit
    - Pour chaque objectif d'audit, une liste de risques/contrôles/tests :
        - Soit issus/inspirés de la base de référence IFACI, et adaptés au contexte d'audit
        - Soit générés par le modèle de langage, sans référence à la base IFACI
    - Une synthèse exécutive
- L'utilisateur est invité à qualifier les différents éléments de la matrice, en particulier ceux générés par le modèle, et ce préalablement à l'export Excel.

4 Impact, risques et mitigation

Utilisateurs concernés : adhérents IFACI et partenaires disposant d'un accès à l'agent Gaia RCM V2.

4.1 Impacts positifs attendus

Le déploiement de l'agent Gaia RCM V2 vise à générer plusieurs impacts positifs, incluant :

- Montée en compétence des utilisateurs : grâce à des réponses structurées, sourcées et contextualisées, Gaia RCM V2 soutient la compréhension et l'appropriation des bonnes pratiques professionnelles.
- Amélioration de la productivité : Gaia RCM V2 permet de libérer du temps pour l'analyse et la validation humaine.

- Support à la qualité et à la conformité : les fonctionnalités de génération, de validation 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 : certaines composantes clés peuvent être absentes ou mal structurées.
- Risque de dépendance excessive : l'utilisation intensive 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 : une partie des réponses s'appuie sur une base de risques/contrôles/tests, elle-même issue d'un corpus documentaire validé (IPPF, normes, documents IFACI), limitant les hallucinations et renforçant la fiabilité.
- Disclaimers : Un disclaimer est présent en début de conversation, dans l'interface de génération RCM et ajouté à chaque export Excel.
- Instructions : Les prompts systèmes sont rédigés de manière à limiter les risques dès la conception. Ils encadrent le comportement de Gaia RCM V2 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 : la génération du contexte d'audit suit un processus structuré et impose une validation finale avant génération de la RCM.

- 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 : L'utilisateur est invité à qualifier les différents éléments de la matrice, en particulier ceux générés par le modèle, et ce préalablement à l'export Excel. Un mécanisme de détection du contenu inapproprié (dans les « prompts » et dans les réponses générées) est implémenté.

5 Confidentialité et protection des données

5.1 Non-entrainement des modèles

Les modèles utilisés sont hébergés en Europe sur infrastructure Azure (AzureOpenAI). Les prompts (questions) et complétions (réponses) ne sont pas utilisés pour améliorer ou entrainer les modèles OpenAI ou Microsoft[2].

5.2 Hébergement

L'infrastructure de Gaia RCM V2 est hébergée sur un environnement Azure propre à l'IFACI en Europe (Azure West Europe) : données, serveurs, services d'inférence LLM.

5.3 Traitement et stockage des données

Gaia RCM V2 (beta) ne stocke ni les questions ni ses réponses complètes[3].
La matrice complète n'est pas stockée, de même que l'executive summary. La conversation (génération du contexte d'audit) et la RCM sont stockées dans le navigateur internet de l'utilisateur, en local.

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* »)

Sont conservés, de façon anonyme à des fins d'amélioration de la plateforme :
- une intention courte sur quelques mots déduite de l'objectif d'audit (par exemple "procurement and inventory audit") ;
- le macro process et l'industrie concernée (par exemple "Deliver Physical Products" et "Manufacturing") ;

[2] https://learn.microsoft.com/en-us/azure/ai-foundry/responsible-ai/openai/data-privacy
[3] A l'exception potentielle du contenu inapproprié, voir « Détection et prévention du contenu inapproprié »

- une version condensée et purgée de toute information sensible des risques/contrôles/tests non sourcés depuis la base IFACI (cellules violettes de la RCM) ;
- les éventuels retours et appréciations (cf ci-dessous).

5.4 Données personnelles et confidentielles

Aucune donnée personnelle (nom, e-mail, …) n'est stockée dans Gaia RCM V2.

Seul un identifiant pseudonymisé, généré à partir des informations de connexion nominatives 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[4].

5.5 Détection et prévention du contenu inapproprié

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 RCM V2 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 RCM V2 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.

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

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.

5.6 Évaluation des réponses

Gaia RCM V2 intègre une fonctionnalité d'évaluation de la qualité des éléments de la RCM par les utilisateurs, sous la forme de boutons *J'aime/Je n'aime pas/A revoir* placés sous chaque cellule de la matrice.
Par ailleurs, au moment de l'export Excel, l'utilisateur est invité à évaluer son expérience : note entre 1 et 5 et commentaire facultatif.
Ces informations sont stockées sur la plateforme lors de l'export Excel, pour analyse statistique et support aux futures évolutions de Gaia RCM V2. Elles ne sont pas conservées si la matrice RCM n'est pas exportée sous Excel.

La phase conversationnelle de génération guidée du contexte d'audit propose une fonction de signalement *j'aime/je n'aime pas*.
Lorsqu'un utilisateur clique sur l'un de ces boutons, la conversation en cours est automatiquement transmise en clair par e-mail au responsable désigné de l'IFACI afin de permettre une revue humaine. Aucune donnée de conversation n'est stockée dans l'infrastructure de Gaia lors de l'utilisation de cette fonctionnalité, seul le mail envoyé contient la conversation.

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 Gaia RCMV2
- Nombre d'utilisateurs uniques
- 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
- Note moyenne
- Commentaires anonymisés
- Performances du système de RAG et du modèle LLM de raisonnement

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

- Utilisation de l'agent Gaia RCM V2, intention associée (résumé sur 5 mots max, par exemple "procurement and inventory audit"), process et industrie associés (listes fermées, par exemple "Deliver Physical Products" et "Manufacturing")
- Consultation d'un document source
- Redirection vers une formation IFACI
- Utilisation des fonctionnalités Like / Dislike / A revoir
- Erreur lors de la génération d'une réponse
- Version condensée et purgée de toute information sensible des risques/contrôles/tests non sourcés depuis la base IFACI (cellules violettes)
- Liste des éléments de risques/contrôles/tests de la base IFACI utilisés pour construire la réponse.

Les prompts, les réponses, le contexte d'audit et la RCM ne sont jamais stockés dans ces traces d'utilisation et aucune donnée personnelle n'est conservée en clair.

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

Les modèles de langage/embedding utilisés par Gaia RCM V2 (gpt-4o, gpt-5.1 et embeddings-3-large) 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.

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-5.1	https://platform.openai.com/docs/models/gpt-5.1	https://cdn.openai.com/pdf/4173ec8d-1229-47db-96de-06d87147e07e/5_1_system_card.pdf
embeddings-3-large	https://openai.com/index/new-embedding-models-and-api-updates	Non applicable (modèle d'embedding)