

Glossaire simple des termes techniques de l'IA

IA générative, agents, alignement, résilience, RAG, IA-native...

Termes fondamentaux

IA générative

IA capable de produire du contenu : texte, image, code, audio, vidéo, idées, plans, etc. Exemples : ChatGPT, Claude, Midjourney, Runway, Suno.

LLM

Large Language Model, ou grand modèle de langage. C'est le type de modèle derrière ChatGPT, Claude, Gemini, Mistral, Llama, etc. Il prédit et génère du texte à partir d'énormes volumes de données.

Modèle multimodal

Modèle capable de traiter plusieurs types d'entrées : texte, image, audio, vidéo, fichiers, parfois écran ou actions.

Prompt

Instruction donnée à une IA. Exemple : "Résume ce texte en 5 points" ou "Agis comme un expert SEO".

Prompt engineering

Art de formuler les bonnes instructions pour obtenir de meilleurs résultats. Les modèles modernes comprennent mieux les demandes simples, mais la structuration reste utile pour les tâches complexes.

Comportement et fiabilité de l'IA

Alignement

Fait de rendre une IA compatible avec les intentions, valeurs, règles et attentes humaines : éviter les réponses dangereuses, reconnaître ses limites, ne pas manipuler, etc.

Hallucination

Quand l'IA invente une information fautive mais la présente comme vraie : étude inexistante, citation inventée, chiffre non vérifié, etc.

Grounding

Fait d'ancrer la réponse de l'IA dans des sources fiables ou dans des données fournies : PDF, base documentaire, résultats web, CRM, documentation interne.

RAG

Retrieval-Augmented Generation. Méthode où l'IA cherche d'abord de l'information dans une base documentaire, puis génère une réponse à partir des documents trouvés.

Contexte

Informations disponibles dans la conversation ou dans les fichiers fournis au modèle. Plus il est riche et bien structuré, plus l'IA peut répondre précisément.

Fenêtre de contexte

Quantité maximale d'informations qu'un modèle peut prendre en compte en une fois.

Mémoire

Capacité d'un système IA à conserver certaines informations entre plusieurs conversations. À distinguer du contexte immédiat.

Business, agents et entreprises IA

Société IA-native

Entreprise conçue dès le départ autour de l'IA, pas seulement une entreprise traditionnelle qui ajoute une fonction IA.

AI-first

Approche où l'IA est pensée comme le cœur de la stratégie produit ou business.

Copilot

Assistant IA qui aide un humain à travailler plus vite, mais sans agir totalement seul.

Agent IA

IA capable de planifier, utiliser des outils, exécuter des actions et poursuivre un objectif, pas seulement répondre en texte.

Agentic AI

IA "agentique", orientée action : elle peut décider d'étapes, appeler des outils, vérifier ses résultats et corriger sa trajectoire.

Agentic workflow

Workflow où plusieurs étapes sont pilotées par une IA ou plusieurs agents : veille, résumé, scoring, tâches, relances, etc.

Agentic OS

Concept d'un système d'exploitation personnel ou professionnel piloté par agents IA, centralisant tâches, documents, mails, calendrier, CRM, objectifs et automatisations.

Résilience, robustesse et gouvernance

Résilience en IA

Oui, le terme existe. Il désigne la capacité d'un système IA à continuer à fonctionner correctement malgré des erreurs, des données imparfaites, des attaques, des changements de contexte ou des pannes.

Robustesse

Capacité d'un modèle à bien fonctionner même quand les données changent, sont imparfaites ou inattendues.

AI safety

Champ qui étudie comment rendre les IA sûres, fiables, contrôlables et non dangereuses.

AI governance

Ensemble des règles, processus et responsabilités pour encadrer l'usage de l'IA dans une organisation.

Données et personnalisation

Dataset

Jeu de données utilisé pour entraîner, tester ou évaluer une IA.

Training data

Données utilisées pour entraîner un modèle.

Fine-tuning

Réentraînement spécialisé d'un modèle sur des données particulières : ton d'une marque, domaine métier, vocabulaire spécifique.

Embeddings

Représentation numérique du sens d'un texte, d'une image ou d'un document. Très utile pour la recherche sémantique.

Vector database

Base de données spécialisée pour stocker des embeddings. Exemples : Pinecone, Weaviate, Chroma, Qdrant.

Données synthétiques

Données générées artificiellement, souvent par IA, pour entraîner ou tester des modèles.

Évaluation et contrôle qualité

Benchmark

Test standardisé pour comparer les performances de plusieurs modèles.

Eval

Évaluation d'un modèle sur une tâche précise : exactitude, hallucinations, suivi de consigne, qualité du code, etc.

Red teaming

Test agressif d'un système IA pour trouver ses failles.

Biais

Tendance d'un modèle à produire des réponses déséquilibrées, injustes ou stéréotypées, souvent à cause des données d'entraînement.

Explicabilité

Capacité à comprendre pourquoi une IA a donné une certaine réponse ou pris une certaine décision.

Types de modèles et déploiement

Modèle propriétaire

Modèle fermé, contrôlé par une entreprise : GPT d'OpenAI, Claude d'Anthropic, Gemini de Google, etc.

Modèle open-weight

Modèle dont les poids sont disponibles, sans être forcément totalement open source.

Modèle open source

Modèle publié avec code, poids et licence permettant une vraie réutilisation ouverte. Le terme est parfois utilisé abusivement en IA.

SLM

Small Language Model : modèle plus petit, moins coûteux, utile pour des tâches ciblées ou locales.

On-device AI

IA qui fonctionne directement sur l'appareil de l'utilisateur : PC, smartphone, objet connecté.

Edge AI

IA exécutée près de la source des données : machine industrielle, caméra, capteur ou appareil local.

Termes stratégiques utiles

Moat IA

Avantage défensif difficile à copier dans une entreprise IA : données propriétaires, distribution, expertise métier, intégration profonde, marque.

Data moat

Avantage concurrentiel basé sur des données uniques.

Workflow automation

Automatisation de processus métiers avec ou sans IA.

Human-in-the-loop

Système où l'IA propose ou exécute partiellement, mais un humain vérifie ou valide.

Orchestration

Organisation de plusieurs outils, modèles ou agents pour accomplir une tâche complète.

Tool use

Capacité d'un modèle à utiliser des outils : navigateur, calendrier, base de données, email, code, API, etc.

Function calling

Mécanisme par lequel un modèle appelle une fonction ou une API de manière structurée.

Termes avancés

Inference

Moment où le modèle produit une réponse à partir d'une entrée. L'entraînement est la phase d'apprentissage ; l'inférence est l'utilisation.

Latence

Temps de réponse d'un système IA.

Coût par token

Coût d'utilisation d'un modèle selon la quantité de texte traitée ou générée.

Token

Morceau de texte utilisé par les modèles. Un mot peut correspondre à un ou plusieurs tokens.

Distillation

Technique pour entraîner un petit modèle à imiter un plus grand modèle.

Mixture of Experts / MoE

Architecture où seuls certains "experts" internes du modèle sont activés selon la demande, pour gagner en efficacité.

Model routing

Choix automatique du bon modèle selon la tâche : petit modèle pour tâche simple, gros modèle pour raisonnement complexe.

Fallback model

Modèle de secours utilisé si le modèle principal échoue, coûte trop cher ou est indisponible.

À maîtriser en priorité

Liste courte

Alignement, RAG, agent IA, agentic workflow, IA-native, grounding, hallucination, human-in-the-loop, orchestration, fine-tuning, embeddings, vector database, AI governance, robustesse, résilience, data moat, model routing.