Glossaire des IA génératives

Emmanuel Desmontils
FST, LS2N, Nantes Université
emmanuel.desmontils@univ-nantes.fr

Version : 1.3

Résumé : Glossaire des principaux termes liés aux IA génératives. Chaque terme est rattaché à un thème ; les définitions s'appuient sur une source citée.

Mots-clés : IA générative, LLM, agents, prompt, hallucination, RGPD, AI Act

Ce glossaire regroupe les termes du cours, par ordre alphabétique, avec leur **thème**. Les définitions s'appuient sur une source citée (voir les références en fin de document) ou, à défaut, sur le contenu du cours.

Thèmes

Données et droit 18 Fondements de l'IA 13 Fonctionnement des modèles 17 Fiabilité 8 Agents 8 Humain et société 8 Environnement 2 Outils 2 Prompt 5 Développement 2

A

Agent

Agents

Entité capable de raisonner et d'agir de façon autonome, en fonction du contexte : elle perçoit son environnement, planifie, utilise des outils (API, moteurs de recherche, bases de données) et itère jusqu'à accomplir sa mission (ou pas).

Agentification

Agents

Processus qui enrichit un LLM ou un chatbot de capacités de perception, de planification et d'action (outils externes, mémoires à long terme, mécanismes de décision), pour passer d'un dialogue réactif à un comportement autonome et itératif.

Agrégateur

Outils

Service en ligne donnant accès à plusieurs modèles de langage (par exemple POE, Duck.ai, OpenRouter, Hugging Face).

Anonymisation et pseudonymisation

Données et droit

Anonymiser, c'est rendre impossible tout lien avec une personne précise. Pseudonymiser ne suffit pas : on peut retrouver la personne.

Apprentissage automatique

Fondements de l'IA

Branche de l'IA dite « numérique » : le système apprend à percevoir et à détecter des corrélations à partir de données (pensée « intuitive et rapide »). Son fonctionnement ne peut pas s'expliquer par introspection.

Apprentissage et inférence
Fonctionnement des modèles

Deux algorithmes à distinguer : l'apprentissage ajuste les paramètres du modèle (descente de gradient) ; l'inférence produit une sortie avec le modèle déjà entraîné (transformer, attention, échantillonnage).

Apprentissage profond
Fondements de l'IA

Apprentissage automatique à base de réseaux de neurones artificiels comportant plusieurs couches de traitement, qui apprennent des représentations des données à plusieurs niveaux d'abstraction. C'est la base technique des systèmes d'IA générative, qui sont donc probabilistes. (LeCun et al., 2015)

Apprentissage supervisé
Fondements de l'IA

Apprentissage à partir d'exemples associant une entrée à la réponse attendue (par exemple un prompt et sa réponse). Le modèle ajuste ses paramètres pour réduire l'écart avec ces réponses.

Atrophie cognitive
Humain et société

Réduction de l'engagement cognitif et de l'effort intellectuel nécessaire pour transformer une information en connaissance. Elle est fortement accentuée par une mauvaise utilisation de l'IA.

Attention et Transformer
Fonctionnement des modèles

Le Transformer (2017) est l'architecture qui sert de base aux LLM. Son mécanisme d'attention pèse l'importance de chaque token du contexte pour prédire le suivant. (Vaswani et al., 2017)

B

Bi-linguisme
Humain et société

Selon D. Andler, nécessité de savoir faire avec les systèmes d'IA (vie professionnelle et sociale) et de savoir faire sans (apprendre à écrire, compter, raisonner ; pouvoir revenir en arrière). (Commission des affaires culturelles, 2024)

Biais
Fiabilité

Écart dans les productions du modèle qui reflète des déséquilibres (genre, origine, âge, handicap…) : par exemple, un « couple » généré par IA presque toujours hétérosexuel, blanc, jeune, mince et sans handicap.

Biais de complaisance (sycophancy)
Humain et société

Biais qui ne vient pas des données mais de l'entraînement : tendance d'une IA à privilégier les réponses qui concordent avec l'opinion de l'utilisateur plutôt que les réponses vraies (sycophancy), possible effet du feedback humain utilisé pour l'entraîner. Elle approuve, rassure et reformule dans le sens de l'utilisateur ; on ose alors moins dire « je ne sais pas » et on supporte moins la contradiction. (Sharma, M. et al., 2024)

Boucle
Agents

Système où le résultat est analysé puis amélioré par rétroaction (par exemple en modifiant le prompt) : l'agent agit, observe, corrige, recommence. Il faut un critère d'arrêt clair, sinon la boucle dérive. (Mermet, 2026)

C

Centre de données (data center) Environnement

Infrastructure de calcul qui fait tourner l'IA : forte consommation d'électricité et d'eau (refroidissement), impact variable selon le mix électrique et la localisation. (International Energy Agency, 2025; Li et al., 2025)

Chaîne de pensée (COT) Prompt

Technique qui demande au modèle d'exposer ses étapes de raisonnement avant de répondre, ce qui améliore les résultats sur les tâches multi-étapes. À l'origine des modèles de raisonnement. (Wei et al., 2022)

Charte d'usage Données et droit

Texte d'un établissement qui fixe les principes d'usage de l'IA générative (l'humain au centre, responsabilité, données, transparence sur l'usage). Exemples : Université d'Orléans (consignes par usage) et Nantes Université (quatre principes). (Nantes Université, 2025; Université d'Orléans, 2024)

Chatbot et moteur de recherche Fiabilité

Un moteur de recherche retrouve et classe des pages existantes, dont les sources sont visibles. Un chatbot d'IA générative produit une réponse nouvelle, suite probable de tokens, sans source visible sauf RAG ou accès au Web : d'où le risque d'hallucination.

Compagnon virtuel Humain et société

IA conçue pour être engageante et disponible à toute heure, au ton chaleureux, sans jugement ni désaccord. Risque : un attachement excessif, au détriment des relations humaines réelles.

D

Deepfake (hypertrucage) Fiabilité

Image, vidéo ou voix générée ou modifiée par IA pour imiter une vraie personne de façon très réaliste. Usages : divertissement, mais aussi manipulation, fraude, harcèlement.

Déontologie Données et droit

Ensemble de règles et de devoirs qui régissent une profession, la conduite de ceux qui l'exercent et leurs rapports avec leurs clients ou le public. (Larousse, 2008)

Détecteur d'IA Fiabilité

Outil censé reconnaître un texte généré par IA. Il est peu fiable et facile à tromper (retouche à la main, changement des paramètres) : ce n'est pas une preuve. On privilégie les preuves de processus (versions, brouillons, jalons). (Massachusetts Institute of Technology, 2026)

Dette cognitive Humain et société

Faible engagement cérébral observé (EEG) lors de la rédaction d'essais avec un LLM : la connectivité est d'autant plus faible que l'aide externe est forte, et le sentiment d'appropriation du texte est faible. Étude à petit échantillon, préprint. (Kosmyna et al., 2025)

Distillation
Fonctionnement des modèles

Entraîner un petit modèle (élève) à imiter les réponses d'un grand modèle (enseignant) : on obtient un modèle plus léger et moins coûteux, au prix d'une certaine perte de qualité. (Hinton et al., 2015)

Données personnelles
Données et droit

Toute information se rapportant à une personne identifiée ou identifiable (nom, téléphone, adresse, numéro de sécurité sociale, informations bancaires…). (Union européenne, 2016; Université d'Orléans, 2024)

Données sensibles
Données et droit

Catégorie particulière de données personnelles (origines raciales ou ethniques, opinions politiques, philosophiques ou religieuses, appartenance syndicale, santé, vie sexuelle…), dont le traitement est en principe interdit sauf exceptions. (Union européenne, 2016)

Droit à l'image
Données et droit

Toute personne a un droit exclusif sur son image : l'image est une donnée personnelle. Un accord écrit est nécessaire pour la prise de vue et la diffusion, et le retrait reste possible (exceptions : lieu public, actualité). (Roubieu, 2024)

Droit d'auteur
Données et droit

Toute œuvre de l'esprit est la propriété de son auteur. Droit moral (contrôle de l'usage et de la citation, non cédable) et droit patrimonial (rémunération, cessible, 70 ans après la mort de l'auteur). La citation d'un court extrait reste permise, avec l'auteur et la référence. (République française, 2026)

E

Effondrement des modèles
Fiabilité

Perte de qualité et de diversité d'un modèle entraîné sur des contenus générés par d'autres IA (ou par lui-même) : propagation des erreurs, des hallucinations et des biais, uniformisation des résultats. (Shumailov et al., 2024)

Embedding
Fonctionnement des modèles

Représentation d'un token par un point (un vecteur) dans un espace : la proximité entre deux points traduit une proximité de sens.

Empreinte écologique de l'IA
Environnement

Impact (électricité, gaz à effet de serre, eau, matières premières) de la construction des modèles et du matériel, et de leur utilisation. Repère : une requête médiane de texte Gemini consomme environ 0,24 Wh d'énergie et 0,26 mL d'eau selon Google. (Elsworth et al., 2025)

Esprit critique
Humain et société

Une IA ne sait pas ce qui est vrai ou faux : elle prédit une suite plausible. Face à une réponse, trois gestes : questionner (d'où vient l'information ?), recouper (avec d'autres sources fiables), assumer (on répond du résultat).

Éthique et cyberéthique

Données et droit

Éthique : partie de la philosophie qui étudie les fondements de la morale, ou ensemble de règles de conduite. Cyberéthique (éthique du numérique) : intégration de critères moraux dans le fonctionnement et l'usage du numérique (médecine, biométrie, IA…). (Germain et al., 2022; Larousse, 2008)

Éthique par délégation

Données et droit

Questions d'éthique soulevées quand on délègue à l'IA : recherche Web, accès au système, interface entre utilisateur et application, agentification des processus.

Explicabilité

Fiabilité

Capacité à expliquer pourquoi un système produit telle sortie. Elle est difficile pour les réseaux de neurones profonds, qui fonctionnent comme des « boîtes noires ». (Barredo Arrieta & others, 2020)

F

Fenêtre de contexte

Fonctionnement des modèles

Quantité de tokens que le modèle peut prendre en compte à la fois (de 2 048 à plus de 10 millions de tokens).

Filigrane textuel

Données et droit

Léger biais statistique dans le choix de certains mots et tournures, qui forme une signature reconnaissable après quelques centaines de mots. Il permet d'estimer, sans certitude, qu'un modèle est intervenu. Il est peu robuste : réécrire le texte avec un autre modèle suffit à le détruire. Base légale du marquage : art. 50 de l'AI Act. (Union européenne, 2024)

Friction cognitive

Humain et société

Effort productif nécessaire pour apprendre, à préserver : l'IA doit aider à penser, pas penser à notre place. Se tourner vers l'IA dès la première difficulté (« cognitive surrender ») donne une illusion d'apprentissage. (Massachusetts Institute of Technology, 2026)

G

GAN (réseau antagoniste génératif)

Fonctionnement des modèles

Deux réseaux s'entraînent l'un contre l'autre : un générateur crée des contenus, un discriminateur essaie de distinguer le faux du vrai. Technique de génération d'images antérieure à la diffusion. (Goodfellow et al., 2014)

H

Hallucination

Fiabilité

Contenu plausible et fluide, mais faux ou non fondé. Elle est factuelle (contradiction d'un fait vérifiable, invention de références) ou de fidélité (consigne non suivie, dérive du contexte). (Huang, L. et al., 2023)

Hiver de l'IA

Fondements de l'IA

Période où les résultats sont en dessous des attentes, d'où désillusion et baisse des financements. (Gefen, 2026)

Humain dans la boucle

Agents

Supervision humaine des résultats et des décisions : vérifier ce qu'on délègue et garder la main sur les décisions à haut risque (cf. AI Act).

I

IA (intelligence artificielle)

Fondements de l'IA

Capacité de machines à utiliser le langage, à former des abstractions et des concepts, à résoudre des problèmes jusque-là réservés aux humains et à s'améliorer. (McCarthy et al., 2006)

IA Act (AI Act)

Données et droit

Règlement européen visant une IA digne de confiance et le respect des droits fondamentaux. C'est une réglementation « produit » : elle dépend de l'usage prévu et non de la technologie. Elle classe les systèmes par niveau de risque (haut risque : par exemple, évaluer les résultats d'apprentissage). Art. 50 : transparence (chatbots signalés, deepfakes étiquetés, marquage lisible par machine). (Union européenne, 2024)

IA générative (SAIg)

Fondements de l'IA

Ensemble d'outils basés sur des modèles de langage génératifs, qui produisent un résultat (texte, image, vidéo…) à partir d'une description (texte, image…). Base technique : apprentissage profond, donc probabiliste.

IA symbolique

Fondements de l'IA

Branche de l'IA (systèmes experts, Web sémantique) fondée sur une pensée « raisonnée et lente » : réfléchir, comprendre, identifier les causalités ; l'introspection est possible.

Intelligence artificielle générale (AGI)

Fondements de l'IA

IA capable de comprendre, d'apprendre et de s'adapter à une grande variété de problèmes, au niveau humain ou au-delà. La notion n'a pas de définition consensuelle. Le problème de l'arrêt exclut une IA omnisciente et infaillible, mais ne démontre pas à lui seul qu'une IA de niveau humain soit impossible (l'argument de Lucas, tiré de Gödel, est très discuté). (Gray, 2026; Lucas, 1961; Turing, 1936)

J

Jailbreak

Données et droit

Prompt conçu pour contourner les garde-fous d'un modèle et obtenir ce qu'il est censé refuser. Il vise le modèle lui-même, alors que l'injection de prompt détourne une application. (Wei et al., 2023)

L

Licence Creative Commons

Données et droit

Licence qui précise à l'avance ce qu'on peut faire d'une œuvre (la reproduire, la modifier, un usage commercial…) en indiquant l'auteur, de CC BY à CC0. Gratuit ou en accès libre ne veut pas dire libre de droit. (Creative Commons, 2026)

LLM et SLM

Fonctionnement des modèles

LLM (Large Language Model) : de 70 à plus de 1 760 milliards de paramètres, grande fenêtre de contexte, très grosse infrastructure. SLM (Small Language Model) : de 200 millions à 20 milliards de paramètres, petite infrastructure, utilisable même sur un PC portable.

M

MCP (Model Context Protocol)

Agents

Standard ouvert pour connecter des applications d'IA à des systèmes externes (sources de données, outils, flux de travail) ; comparé à un port USB-C pour l'IA. (Model Context Protocol, 2026)

Modèle de diffusion

Fonctionnement des modèles

Modèle génératif d'images : appris en ajoutant progressivement du bruit à des images, il génère en retirant le bruit pas à pas, guidé par une description. (Ho et al., 2020)

Modèle de fondation

Fonctionnement des modèles

Modèle entraîné sur une sélection de téraoctets de données. Il est ensuite spécialisé (pré-prompt qui fixe les règles d'échange), instruit (apprentissage supervisé) puis finalisé (RLHF).

Modèle de raisonnement

Fonctionnement des modèles

Modèle qui génère d'abord une suite de « pensées » (chaîne de raisonnement, COT) avant sa réponse. Utile pour les tâches multi-étapes ; plus lent et plus coûteux en énergie.

Modèle multimodal

Fonctionnement des modèles

Modèle qui traite et produit plusieurs types de données (texte, image, audio, vidéo). Exemple : un modèle audio vers audio qui prend en compte l'intonation, le rythme et les interruptions. (OpenAI, 2024)

Multi-agents (système)

Agents

Un orchestrateur répartit le travail entre des sous-agents spécialisés qui travaillent en parallèle, chacun avec son propre contexte. Adapté à la recherche large ; environ 15 fois plus de tokens qu'un chat. (Hadfield et al., 2025)

O

Open Education (REL)

Données et droit

Mouvement qui promeut des ressources éducatives libres (REL) : documents d'enseignement, d'apprentissage ou de recherche réutilisables, souvent sous licence ouverte. (UNESCO, 2019)

P

Paramètres (taille d'un modèle) Fonctionnement des modèles

Nombre de valeurs gérées par le modèle.

Poids ouverts et open source Outils

Un modèle à poids ouverts a ses poids téléchargeables et peut tourner en local, mais on n'en connaît ni le code ni les données. Un modèle open source (au sens de l'OSI) publie aussi le code complet d'entraînement et d'exécution et des informations détaillées sur les données d'entraînement, pas forcément les données elles-mêmes. Poids ouverts ne veut pas dire open source. (Mermet, 2026; Open Source Initiative, 2024)

Pré-prompt Prompt

Texte ajouté automatiquement au prompt de l'utilisateur pour guider le LLM et fixer les règles d'échange (modèle « spécialisé »). (Mermet, 2026)

Principes éthiques de l'ISO Données et droit

Sept principes pour une IA responsable : équité, transparence, absence de malveillance, responsabilité, respect de la vie privée, robustesse, inclusivité. (International Standard Organisation (ISO), 2025)

Problème de l'arrêt Fondements de l'IA

Résultat de Turing (1936) : il n'existe aucun algorithme qui décide, pour tout programme et toute entrée, si le programme s'arrête ou boucle indéfiniment. C'est une limite de tout calcul, donc de toute IA. (Turing, 1936)

Prompt Prompt

Instruction donnée au modèle (prompt système, prompt utilisateur). Un cadre classique : persona, tâche, contexte, format. (Google Workspace, 2024)

Prompt injection Données et droit

Entrée qui modifie le comportement du modèle de façon non voulue, même si elle est invisible pour un humain. Dans sa forme indirecte, la consigne est cachée dans un contenu lu par l'IA (page web, document, mail) : le modèle distingue mal les données des instructions. Plus grave avec un agent, qui dispose d'outils. (Greshake et al., 2023; OWASP Gen AI Security Project, 2025)

Q

Quantification Fonctionnement des modèles

Réduction de la précision numérique des paramètres d'un modèle (par exemple 16 bits vers 4 ou 8 bits) pour diminuer sa taille et sa consommation de mémoire, au prix d'une perte de précision. (Jacob et al., 2018)

R

RAG (génération à enrichissement contextuel) Fonctionnement des modèles

Technique qui apporte au modèle du contenu supplémentaire (documents retrouvés) pour qu'il réponde à partir de sources. Dans l'article d'origine, elle associe la mémoire du modèle (ses paramètres) à une mémoire externe consultée par un module de recherche. (Gao et al., 2023; Lewis et al., 2020)

Réseau de neurones Fondements de l'IA

Modèle formé de neurones artificiels reliés entre eux : chacun combine ses entrées avec des poids, puis transmet un signal. Les poids sont ajustés pendant l'apprentissage (idée issue du perceptron, 1958). (LeCun et al., 2015; Rosenblatt, 1958)

Résumé IA dans les moteurs de recherche Fiabilité

Réponse générée par IA en tête des résultats d'un moteur. Les utilisateurs cliquent moins sur les sources, d'où un risque de baisse de visites et de revenus pour les sites de référence (presse, savoirs). (Chapekis & Lieb, 2025)

RGPD Données et droit

Règlement Général sur la Protection des Données (UE, 2016, applicable depuis 2018). Informer les personnes et, selon le cas, obtenir leur accord ; avec une IA, ce qu'on envoie peut être enregistré ou réutilisé. (Union européenne, 2016)

RLHF Fonctionnement des modèles

Apprentissage par renforcement avec feedback humain : des humains classent les réponses du modèle à des prompts, selon des consignes, pour orienter le modèle final. Il suit un fine-tuning supervisé sur des démonstrations humaines. (Ouyang, L. et al., 2022)

S

Simulation multi-agents Agents

Modélisation d'acteurs (par exemple en géographie ou en aménagement du littoral) qui n'utilise pas forcément de LLM. À ne pas confondre avec un système multi-agents à base de LLM.

Six piliers de lecture de l'IAg Fondements de l'IA

Grille d'analyse d'un cas ou d'un discours : les quatre concepts de Dowek (algorithme, machine, langage, information), auxquels s'ajoutent l'humain et la société. (Dowek, 2011)

Sortie contrainte Prompt

Format de réponse imposé au modèle (schéma JSON, appel de fonction, grammaire) : format garanti, intégration directe dans un programme, mais richesse parfois bridée.

Souveraineté numérique Données et droit

Maîtrise de où sont les données, du droit qui s'applique au fournisseur et de qui contrôle l'infrastructure. Pistes : outils institutionnels ou locaux.

T

Techno-solutionnisme

Humain et société

Croyance (critiquée par E. Morozov) qu'il existe une solution technique à chaque problème complexe, au risque d'ignorer leurs dimensions sociales, politiques, culturelles ou écologiques et de remplacer la compréhension par la foi en la technologie. (Morozov, 2013)

Température, top-p, top-k

Fonctionnement des modèles

Réglages du choix du prochain token. Température : degré d'aléatoire (déterminisme contre « créativité » ; reproductibilité non garantie, même à 0). Top-p : garde le plus petit ensemble de tokens dont la probabilité cumulée atteint p. Top-k : ne garde que les k tokens les plus probables. (He & Thinking Machines Lab, 2025)

Test de Turing

Fondements de l'IA

Test proposé par A. Turing (1950, « jeu de l'imitation ») : une machine est jugée « intelligente » si un interrogateur, par écrit, ne la distingue pas d'un humain. (Turing, 1950)

Théorème de Rice

Fondements de l'IA

Extension du problème de l'arrêt (1953) : toute propriété non triviale du comportement d'un programme est indécidable en général. Il n'existe donc pas de vérificateur universel qui garantisse qu'un programme (ou une IA) est correct ou sûr dans toutes les situations. (Rice, 1953)

Token

Fonctionnement des modèles

Unité de texte traitée par le modèle, qui produit sa réponse comme une suite probable de tokens, un token après l'autre.

V

Vibe coding

Développement

Programmer en décrivant ce qu'on veut en langage naturel à une IA et en acceptant son code sans le relire en détail. Pratique rapide mais risquée (erreurs, failles, code non maîtrisé). (Karpathy, 2025)

Vulnérabilité de code

Développement

Faille de sécurité dans un programme. Dans une étude de 2021, environ 40 % des programmes générés par Copilot en présentaient. (Pearce et al., 2022)

W

Workflow (orchestrateur)

Agents

Un orchestrateur enchaîne plusieurs appels de LLM avec des prompts spécialisés ; le flux d'exécution est entièrement prédéfini, le comportement est déterministe et prévisible. (Schluntz & Zhang, 2024)

Z

Zero-shot et few-shot Prompt

Zero-shot : on pose la tâche sans exemple. Few-shot : on donne dans le prompt un ou quelques exemples du format attendu, ce qui est plus fiable qu'une simple description. (Brown et al., 2020)

Références

Barredo Arrieta, A., & others. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Brown, T. B., Mann, B., Ryder, N., & others. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 1877-1901.

Chapekis, A., & Lieb, A. (2025,). *Google users are less likely to click on links when an AI summary appears in the results*. https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/

Commission des affaires culturelles. (2024, février). *École et Intelligence Artificielle* [Vidéo d'une Table ronde]. https://videos.assemblee-nationale.fr/video.14614052_65c33db21467b?timecode=658320

Creative Commons. (2026). *Licences Creative Commons*. https://creativecommons.org/share-your-work/cclicenses/

Dowek, G. (2011). Les quatre concepts de l'informatique. *Sciences et technologies de l'information et de la communication en milieu éducatif : analyse de pratiques et enjeux didactiques*, 21-29. https://edutice.archives-ouvertes.fr/edutice-00676169

Elsworth, C., Huang, K., Patterson, D., Schneider, I., Sedivy, R., Goodman, S., Townsend, B., Ranganathan, P., Dean, J., Vahdat, A., Gomes, B., & Manyika, J. (2025). Measuring the environmental impact of delivering AI at Google Scale. *arXiv preprint arXiv:2508.15734*. https://arxiv.org/abs/2508.15734

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, *2*(1). https://arxiv.org/abs/2312.10997

Gefen, A. (éd.). (2026). *Histoire culturelle de l'IA*. CNRS Éditions. https://www.fabula.org/actualites/135499/une-histoire-culturelle-de-l-intelligence-artificielle-27-08.html

Germain, E., Kirchner, C., & Tessier, C. (2022). Regards croisés sur deux années de réflexion du CNPEN. In E. Germain, C. Kirchner, & C. Tessier (éds.), *Pour une éthique du numérique* (p. 11-16). puf. https://www.puf.com/pour-une-ethique-du-numerique

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, 2672-2680.

Google Workspace. (2024). *Prompting guide 101: a quick-start handbook for effective prompts* [Technical report]. https://services.google.com/fh/files/misc/gemini_for_workspace_prompt_guide_october_2024_digital_final.pdf

Gray, P. (2026, septembre). *Alan Turing identified a logical boundary that keeps AI from knowing everything*. https://www.futura-sciences.com/en/the-shocking-flaw-alan-turing-saw-why-ai-will-never-surpass-us_38914/

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023, février). *Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection*. https://arxiv.org/abs/2302.12173

Hadfield, J., Zhang, B., Lien, K., Scholz, F., Fox, J., & Ford, D. (2025, juin). *How we built our multi-agent research system*. https://www.anthropic.com/engineering/multi-agent-research-system

He, H., & Thinking Machines Lab. (2025, septembre). *Defeating Nondeterminism in LLM Inference*. https://thinkingmachines.ai/blog/defeating-nondeterminism-in-llm-inference/

Hinton, G., Vinyals, O., & Dean, J. (2015, mars). *Distilling the Knowledge in a Neural Network*.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 6840-6851.

Huang, L. et al. (2023, novembre). *A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions*. https://arxiv.org/abs/2311.05232

International Energy Agency. (2025). *Energy and AI* [Technical report]. https://www.iea.org/reports/energy-and-ai

International Standard Organisation (ISO) (éd.). (2025, juillet). *Bâtir une IA responsable : comment aborder le débat sur l'éthique de l'IA ?*. https://www.iso.org/fr/intelligence-artificielle/ethique-ia-responsable

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2018). Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2704-2713.

Karpathy, A. (2025, février). *There's a new kind of coding I call ``vibe coding''*. https://x.com/karpathy/status/1886192184808149383

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. *arXiv preprint arXiv:2506.08872*. https://arxiv.org/abs/2506.08872

Larousse. (2008). *Le Petit Larousse illustré 2009* (p. 1889). Éditions Larousse.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. https://doi.org/10.1038/nature14539

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020, décembre). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. https://arxiv.org/abs/2005.11401

Li, P., Yang, J., Islam, M. A., & Ren, S. (2025). Making AI Less ``Thirsty'': Uncovering and Addressing the Secret Water Footprint of AI Models. *arXiv preprint arXiv:2304.03271 (accepté par Communications of the ACM)*. https://arxiv.org/abs/2304.03271

Lucas, J. R. (1961). Minds, Machines and Gödel. *Philosophy*, *36*(137), 112-127. https://doi.org/10.1017/S0031819100024591

Massachusetts Institute of Technology. (2026). *Final Report of the Ad Hoc Committee on AI Use in Teaching, Learning, and Research Training* [Rapport]. https://aiandeducation.mit.edu/report/

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence: August 31, 1955. *AI magazine*, *27*(4), 12-14. https://doi.org/10.1609/aimag.v27i4.1904

Mermet, B. (2026). *Formation IA Formateurs NSI & SNT : glossaire*. https://formation1-1e8393.gitlab.io/310-glossaire.html

Model Context Protocol. (2026). *What is the Model Context Protocol (MCP)?*. https://modelcontextprotocol.io/

Morozov, E. (2013). *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist*. PublicAffairs.

Nantes Université. (2025). *4 principes institutionnels pour des usages responsables et inclusifs des IAGen* [Technical report]. https://www.univ-nantes.fr/universite/vision-strategie-et-grands-projets/nantes-universite-sengage-pour-une-utilisation-responsable-et-inclusive-des-iagen

Open Source Initiative. (2024, octobre). *The Open Source AI Definition – 1.0*. https://opensource.org/ai/open-source-ai-definition

OpenAI. (2024, mai). *Hello GPT-4o*.

Ouyang, L. et al. (2022, novembre). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*. https://arxiv.org/abs/2203.02155

OWASP Gen AI Security Project. (2025). *LLM01:2025 Prompt Injection (OWASP Top 10 for LLM Applications 2025)*. https://genai.owasp.org/llmrisk/llm01-prompt-injection/

Pearce, H., Ahmad, B., Tan, B., Dolan-Gavitt, B., & Karri, R. (2022, mai). Asleep at the Keyboard? Assessing the Security of GitHub Copilot's Code Contributions. *IEEE Symposium on Security and Privacy (S&P)*. https://arxiv.org/abs/2108.09293

Rice, H. G. (1953). Classes of recursively enumerable sets and their decision problems. *Transactions of the American Mathematical Society*, *74*(2), 358-366. https://doi.org/10.1090/S0002-9947-1953-0053041-6

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386-408. https://doi.org/10.1037/h0042519

Roubieu, S. (2024). *Petit livret de déontologie numérique* [Cours].

République française. (2026). *Code de la propriété intellectuelle, art. L111-1 et suivants*. https://www.legifrance.gouv.fr/codes/id/LEGITEXT000006069414/

Schluntz, E., & Zhang, B. (2024, décembre). *Building effective agents*. https://www.anthropic.com/research/building-effective-agents

Sharma, M. et al. (2024, mai). Towards Understanding Sycophancy in Language Models. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2310.13548

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755-759. https://doi.org/10.1038/s41586-024-07566-y

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *1*, 230-265. https://doi.org/10.1112/plms/s2-42.1.230

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

UNESCO. (2019). *Recommandation sur les ressources éducatives libres (REL)* [Technical report]. https://www.unesco.org/fr/communication-information/oer-recommendation

Union européenne. (2016, avril). *Règlement (UE) 2016/679 du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel (RGPD)*. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Union européenne. (2024, juin). *Règlement (UE) 2024/1689 du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Université d'Orléans. (2024). *Charte sur l'utilisation de l'Intelligence Artificielle au sein de l'Université d'Orléans* [Technical report]. https://www.univ-orleans.fr/upload/public/2024-06/dircom_2024_charte_IA_UO_R.pdf

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 5998-6008.

Wei, A., Haghtalab, N., & Steinhardt, J. (2023, décembre). Jailbroken: How Does LLM Safety Training Fail?. *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022, novembre). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*.

Ce document est distribué selon les termes de la licence **Creative Commons** CC BY. Vous pouvez trouver le texte de la licence à l'adresse suivante : https://creativecommons.org/licenses/by/4.0/

Il utilise la police **Atkinson Hyperlegible Next** (Braille Institute of America, Inc.) afin d'améliorer la lisibilité.

ASSISTÉ PAR L'IA