RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed Lamine Debaghine Setif 2

Faculté de Droit et Sciences Politiques

Département des Sciences Politiques

POLYCOPIÉ DE COURS

L'INTELLIGENCE ARTIFICIELLE

الذكاء الاصطناعي

Cours destiné aux étudiants de Sciences Politiques

UE Découverte (الاستكشافية)

Coefficient 02 - Crédits 02

Réalisé par

Dr. BOUNOUNI Mahdi

Maître de Conférences « A » en Informatique

Faculté de Droit et des Sciences Politiques

Département des Sciences Politiques

Année universitaire 2025 / 2026

FICHE SIGNALÉTIQUE DU MODULE

Élément	Valeur
Intitulé (FR)	Intelligence Artificielle
Intitulé (AR)	الذكاء الاصطناعي
UE	Découverte الاستكشافية
Coefficient / Crédits	02 / 02
Volume horaire	42 h (Cours magistral + TD)
Public	Licence 2 en Sciences Politiques
Prérequis	Notions de base en informatique et bureautique
Langue	Français (terminologie bilingue FR / AR)
Enseignant	Dr. BOUNOUNI Mahdi, MCA Informatique
Année universitaire	2025 / 2026

Objectifs généraux du module

Ce module poursuit quatre objectifs complémentaires. En premier lieu, il ambitionne de doter les étudiants d'une culture technique sérieuse sur l'intelligence artificielle : ses mécanismes, ses algorithmes, ses données sans exiger de bagage en mathématiques avancées. En deuxième lieu, il vise à clarifier les concepts techniques en les présentant de manière progressive, avec des analogies accessibles et des exemples du quotidien. En troisième lieu, il met en lumière les dimensions éthiques, juridiques et politiques de l'IA, non comme un développement séparé, mais comme la conséquence naturelle de choix techniques. Enfin, il prépare le futur diplômé à comprendre, questionner et contribuer aux débats publics sur une technologie qui redessine la gouvernance, la diplomatie et la société.

AVANT-PROPOS

Ce polycopié est né d'un constat simple, fait en salle de cours : les étudiants en sciences politiques s'intéressent sincèrement à l'intelligence artificielle, mais la plupart des ressources disponibles les placent face à un choix inconfortable. Soit elles noient le lecteur dans des formules mathématiques qui présupposent une formation en informatique, soit elles survendent la dimension politique au détriment d'une compréhension réelle de ce que la technologie fait et de comment elle le fait.

Ce cours prend un autre chemin. Il part du principe qu'un étudiant en sciences politiques n'a pas besoin de savoir programmer pour comprendre comment un algorithme de recommandation façonne l'opinion publique, comment un modèle de langage génère une propagande politique convaincante, ou pourquoi la qualité des données d'entraînement détermine la qualité et l'équité des décisions automatisées. Mais il a besoin de comprendre ces mécanismes avec suffisamment de précision pour ne pas se laisser impressionner par le jargon ni abuser par les discours marketing.

La structure du cours reflète cet équilibre. Les premiers chapitres posent les bases techniques, ce qu'est réellement un algorithme d'apprentissage, comment un réseau de neurones traite une image ou un texte, pourquoi un grand modèle de langage comme ChatGPT produit parfois des erreurs avec une confiance désarmante. Ces bases ne sont pas facultatives : elles conditionnent la qualité du regard critique que le politiste portera ensuite sur les usages, les risques et les régulations. Les chapitres suivants appliquent ces fondements aux grands enjeux de la gouvernance, de la désinformation, de la surveillance et des relations internationales, non comme des développements purement théoriques, mais comme des analyses ancrées dans des cas documentés.

Je dois signaler une particularité de ce polycopié qui le distingue de beaucoup d'autres : j'ai choisi de décrire précisément les schémas et les visuels pédagogiques que l'enseignant doit dessiner ou projeter en cours, plutôt que de les annoncer vaguement. Un schéma décrit avec précision : blocs, flèches, couleurs, disposition : peut être reproduit au tableau en cinq minutes et transmis bien plus efficacement qu'une illustration floue.

Enfin, toutes les références bibliographiques de ce polycopié sont vérifiées. Les DOI sont réels, les ouvrages existent, les rapports institutionnels sont accessibles aux adresses indiquées. Je vous encourage à consulter les sources en priorité pour les sujets qui vous touchent le plus : la littérature sur l'IA est vaste, vivante et souvent passionnante.

TABLE DES MATIÈRES

LISTE DES ABRÉVIATIONS

Cette liste regroupe les principales abréviations utilisées dans le polycopié afin de faciliter la lecture des chapitres techniques.

Abréviation	Signification	Utilisation dans le polycopié
IA	Intelligence artificielle	Terme central du polycopié.
AI	Artificial Intelligence	Équivalent anglais du terme IA.
ANI	Artificial Narrow Intelligence	IA étroite, spécialisée dans une tâche.
AGI	Artificial General Intelligence	IA générale, encore non réalisée.
ASI	Artificial Superintelligence	Superintelligence artificielle, notion prospective.
ML	Machine Learning	Apprentissage automatique à partir de données.
DL	Deep Learning	Apprentissage profond fondé sur des réseaux neuronaux.
NLP	Natural Language Processing	Traitement automatique du langage naturel.
TALN	Traitement automatique du langage naturel	Domaine de l'IA appliqué au texte et à la parole.
LLM	Large Language Model	Grand modèle de langage, ex. ChatGPT, Claude, Gemini.
CNN	Convolutional Neural Network	Réseau neuronal utilisé surtout en vision par ordinateur.
RNN	Recurrent Neural Network	Réseau neuronal adapté aux séquences.
SVM	Support Vector Machine	Algorithme de classification supervisée.
k-NN	k-Nearest Neighbors	Méthode des k plus proches voisins.
RAG	Retrieval-Augmented Generation	Génération augmentée par récupération documentaire.
OCR	Optical Character Recognition	Reconnaissance optique de caractères.
API	Application Programming Interface	Interface permettant à deux logiciels de communiquer.
GPU	Graphics Processing Unit	Processeur graphique utilisé pour entraîner les modèles.
CPU	Central Processing Unit	Processeur central de l'ordinateur.

Abréviation	Signification	Utilisation dans le polycopié
IoT	Internet of Things	Internet des objets connectés.
RPA	Robotic Process Automation	Automatisation de tâches administratives répétitives.
XAI	Explainable Artificial Intelligence	IA explicable, utile pour comprendre les décisions.
RGPD	Règlement général sur la protection des données	Cadre européen de protection des données.
OCDE	Organisation de coopération et de développement économiques	Source de principes internationaux sur l'IA.
UNESCO	Organisation des Nations Unies pour l'éducation, la science et la culture	Source de recommandations éthiques sur l'IA.
NIST	National Institute of Standards and Technology	Organisme américain de référence technique.
IEEE	Institute of Electrical and Electronics Engineers	Organisation scientifique et technique.
ACM	Association for Computing Machinery	Association scientifique en informatique.
DOI	Digital Object Identifier	Identifiant numérique d'un article ou document scientifique.

INTRODUCTION GÉNÉRALE

Comprendre l'IA pour agir dans un monde algorithmique

Il y a quelques années encore, l'intelligence artificielle était perçue comme une spécialité réservée aux ingénieurs et aux chercheurs en informatique. Elle évoquait des robots de science-fiction, des superordinateurs qui battaient les champions d'échecs, et des travaux académiques inaccessibles au commun des mortels. Cette image ne suffit plus aujourd'hui. L'IA est partout : dans le moteur de recherche qui complète vos phrases, dans le fil d'actualité qui décide de ce que vous lisez, dans les filtres qui modèrent les contenus en ligne, dans les administrations qui traitent vos dossiers, et dans les systèmes militaires qui guident des drones à l'autre bout du monde.

Pour un étudiant en sciences politiques, cette présence croissante impose une compétence nouvelle. Il ne s'agit plus d'une curiosité technologique optionnelle. Comprendre l'IA, pas nécessairement à un niveau d'ingénierie, mais avec suffisamment de profondeur pour distinguer la réalité de la rhétorique : est devenu une compétence civique et professionnelle indispensable. Comment analyser sérieusement une politique de modération des contenus sans savoir ce qu'est un classifieur ? Comment évaluer les risques d'un système d'aide à la décision judiciaire sans comprendre ce qu'est un biais algorithmique ? Comment peser dans le débat sur la régulation de l'IA sans connaître la différence entre un modèle discriminatif et un modèle génératif ?

L'architecture du cours

Le cours est organisé en trois axes complémentaires, conçus pour progresser du général au spécifique, et du technique au politique.

Le premier axe couvre les fondements techniques. Six chapitres y traitent successivement de la définition de l'IA, du rôle des données et des algorithmes, du Machine Learning et du Deep Learning, des réseaux de neurones, du traitement automatique du langage et de l'IA générative, et enfin de la vision par ordinateur, des systèmes de recommandation et des chatbots. L'objectif n'est pas de former des programmeurs : c'est de donner aux étudiants les images mentales justes, les analogies pertinentes et le vocabulaire précis pour comprendre et critiquer ces technologies.

Le deuxième axe aborde les limites, les risques et les enjeux éthiques. Deux chapitres examinent les erreurs et hallucinations des modèles, les biais algorithmiques, les questions de cybersécurité et d'automatisation, et la notion de souveraineté numérique. Ces chapitres constituent le pont entre la technique et le politique.

Le troisième axe applique les acquis des deux premiers aux domaines classiques des sciences politiques : la gouvernance et l'administration publique, la lutte contre la désinformation et la protection des processus démocratiques, les relations internationales et la géopolitique de l'IA. Chaque application est illustrée par des exemples documentés, des études de cas précises et des références vérifiables.

Remarque méthodologique

L'intelligence artificielle est un domaine dont l'évolution se mesure en mois, pas en décennies. Certains exemples ou performances mentionnés dans ce polycopié seront peut-être dépassés au moment de votre lecture. Ce qui ne change pas, en revanche, ce sont les principes fondamentaux : comment un réseau de neurones apprend, pourquoi les données de mauvaise qualité produisent des modèles biaisés, comment un texte génératif est produit mot après mot. Ces principes sont votre boussole. Utilisez-les pour évaluer chaque nouveauté qui surgira dans le domaine pendant et après votre formation.

AXE 1

FONDEMENTS TECHNIQUES

DE L'INTELLIGENCE ARTIFICIELLE

أسس الذكاء الاصطناعي التقنية

Cet axe pose les fondations indispensables. Avant de s'interroger sur les effets politiques ou éthiques de l'intelligence artificielle, il faut comprendre ce qu'elle est réellement et ce qu'elle n'est pas. Six chapitres se succèdent, chacun ajoutant une couche de compréhension sur la précédente : définition et histoire, données et algorithmes, apprentissage automatique, réseaux de neurones, traitement du langage et IA générative, puis vision par ordinateur et systèmes de recommandation. À l'issue de cet axe, vous disposerez des images mentales et du vocabulaire nécessaires pour analyser n'importe quel usage de l'IA sans être impressionné par le jargon.

Chapitre 1 : Qu'est-ce que l'intelligence artificielle ?

Objectifs pédagogiques

- Définir l'intelligence artificielle de manière rigoureuse et accessible.
- Distinguer les trois niveaux d'IA : étroite, générale, superintelligence.
- Retracer les grandes étapes historiques de la discipline.
- Identifier les composantes fondamentales d'un système d'IA.
- Démystifier les confusions entre IA, robotique, automatisation et programmation classique.

1.1 Définir l'IA : une notion plus précise qu'on ne le croit

Commençons par une question simple : savez-vous ce qui distingue un logiciel de calcul d'impôts d'un système d'intelligence artificielle ? Les deux traitent des données et produisent des résultats automatiquement. Pourtant, ils sont fondamentalement différents. Le logiciel fiscal suit des règles écrites mot à mot par un programmeur humain : si le revenu dépasse tel seuil, appliquer tel taux. Rien ne surprend, rien n'« apprend ». Un système d'IA, lui, est capable de dégager ses propres règles à partir d'exemples sans qu'un humain ait eu à les formuler explicitement.

C'est cette capacité à apprendre, à extraire des régularités et à généraliser à partir de données qui définit le cœur de l'intelligence artificielle moderne. Voici la définition retenue par l'OCDE, la plus citée dans les textes officiels [1] :

> *« Un système d'IA est un système automatisé qui, pour un ensemble donné d'objectifs définis par des humains, émet des prédictions, des recommandations ou des décisions influençant des environnements réels ou virtuels. »* : OCDE, Principes sur l'IA, 2019 (révisés 2024) [1]

Cette définition mérite d'être décortiquée. Elle insiste sur trois points : la machine poursuit des objectifs fixés par des humains (elle n'est pas autonome dans ses buts), elle produit des sorties qui ne sont pas entièrement prévisibles à l'avance (prédictions, recommandations), et ces sorties agissent sur le monde réel. Ce dernier point est crucial pour le politiste : une IA qui décide si un demandeur de logement social est prioritaire n'est pas un simple calcul : c'est un acteur politique au sens fonctionnel.

1.1.1 L'IA n'est pas de la magie : c'est des mathématiques appliquées à des données

L'une des confusions les plus fréquentes consiste à attribuer à l'IA une forme d'intelligence mystérieuse, proche de celle du cerveau humain. Cette image est trompeuse. Un système d'IA est, dans son essence, une fonction mathématique complexe qui transforme des entrées (données) en sorties (résultats). La complexité de cette fonction peut être immense : un grand modèle de langage comme GPT-4 contient plus de mille milliards de paramètres, mais il reste une fonction, calculée par des opérations algébriques, sur du matériel physique qui consomme de l'électricité.

ANALOGIE : Analogie du cuisinier

Imaginez un cuisinier qui, après avoir préparé des milliers de plats et observé les réactions de ses clients, développe une intuition pour doser les épices sans recette. Il ne calcule pas : il a internalisé des régularités. Un modèle d'IA fait quelque chose de similaire, mais avec des nombres. Il observe des millions d'exemples (des plats et leurs notes), ajuste ses paramètres internes (ses dosages numériques), et finit par prédire correctement ce que des clients apprécieront sans avoir jamais eu de recette explicite. La différence fondamentale : le cuisinier comprend le goût. L'IA, non.

1.2 La classification tripartite : étroite, générale, superintelligence

La littérature distingue trois niveaux d'IA selon leur étendue de capacités [2]. Maîtriser cette classification est indispensable pour lire un article spécialisé ou évaluer une annonce technologique.

Type	Désignation (AR)	Capacités	Statut actuel	Exemples
IA étroite (ANI)	ذكاء ضيق	Une seule tâche ou domaine restreint	Existe : largement déployée	ChatGPT, AlphaGo, reconnaissance faciale, filtres anti-spam
IA générale (AGI)	ذكاء عام	Toutes tâches cognitives humaines	N'existe pas : objectif de recherche	Aucun système réel à ce jour
Superintelligence (ASI)	ذكاء فائق	Dépasse l'humain dans tous les domaines	Hypothétique : prospective	Concept philosophique et stratégique

Tableau 1.1 : Les trois niveaux d'IA. Source : classification adaptée de [2].

Attention aux annonces spectaculaires

En 2023 et 2024, plusieurs dirigeants d'entreprises technologiques ont affirmé que l'AGI était « imminente » ou déjà atteinte. Ces déclarations sont à lire avec prudence : elles servent souvent à attirer des investissements ou à influencer les débats réglementaires. En 2026, aucun système d'IA générale n'existe. Tout ce qui est déployé dans le monde réel, y compris les modèles les plus avancés : relève de l'IA étroite, parfois très sophistiquée, mais toujours limitée à un domaine.

1.3 Brève histoire : cinq ruptures qui ont tout changé

L'histoire de l'IA n'est pas un progrès linéaire. C'est une suite d'enthousiasmes, de désillusions et de rebonds qui aident à comprendre pourquoi la discipline ressemble aujourd'hui à ce qu'elle est.

1.3.1 1956 : La naissance officielle : la conférence de Dartmouth

C'est lors d'un séminaire d'été à Dartmouth College, aux États-Unis, que le terme « artificial intelligence » est officiellement adopté par John McCarthy, Marvin Minsky et leurs collègues [3]. L'ambition affichée est démesurée : reproduire l'intelligence humaine dans une machine en quelques années. Cette ambition nourrira des décennies de recherche productive et de déceptions cuisantes.

La première approche dominante, dite symbolique, consiste à coder explicitement des règles logiques. Les systèmes experts des années 1970-1980 en sont l'exemple le plus abouti : ils codifient le savoir d'un médecin ou d'un ingénieur sous forme de milliers de règles SI/ALORS. Ils fonctionnent bien dans des domaines restreints, mais s'effondrent dès que la réalité dépasse les règles prévues.

1.3.2 1956-1990 : Les hivers de l'IA

Deux périodes de désillusion majeures, appelées « hivers de l'IA », frappent la discipline. Le premier survient dans les années 1970 : les promesses ne sont pas tenues, les financements se tarissent. Le second frappe dans les années 1980-1990, au moment où les systèmes experts montrent leurs limites. Ces hivers ne sont pas des échecs totaux : la recherche continue en souterrain, mais ils expliquent le scepticisme que certains chercheurs gardent encore face aux annonces enthousiastes.

1.3.3 2012 : La révolution de l'apprentissage profond

L'année 2012 marque un vrai tournant. Lors du concours ImageNet de reconnaissance d'images, l'équipe de Geoffrey Hinton présente AlexNet, un réseau de neurones profond entraîné sur des processeurs graphiques (GPU) [4]. Le taux d'erreur chute brutalement de 26 % à 15 %, là où toutes les autres approches stagnaient. En quelques mois, l'apprentissage profond devient le nouveau paradigme dominant.

Ce succès repose sur trois ingrédients qui se sont combinés au bon moment des données massives rendues disponibles par Internet, des GPU devenus suffisamment puissants pour entraîner des réseaux profonds, et des algorithmes d'optimisation améliorés. Aucun des trois n'est une nouveauté en 2012, mais leur combinaison produit un effet de seuil spectaculaire.

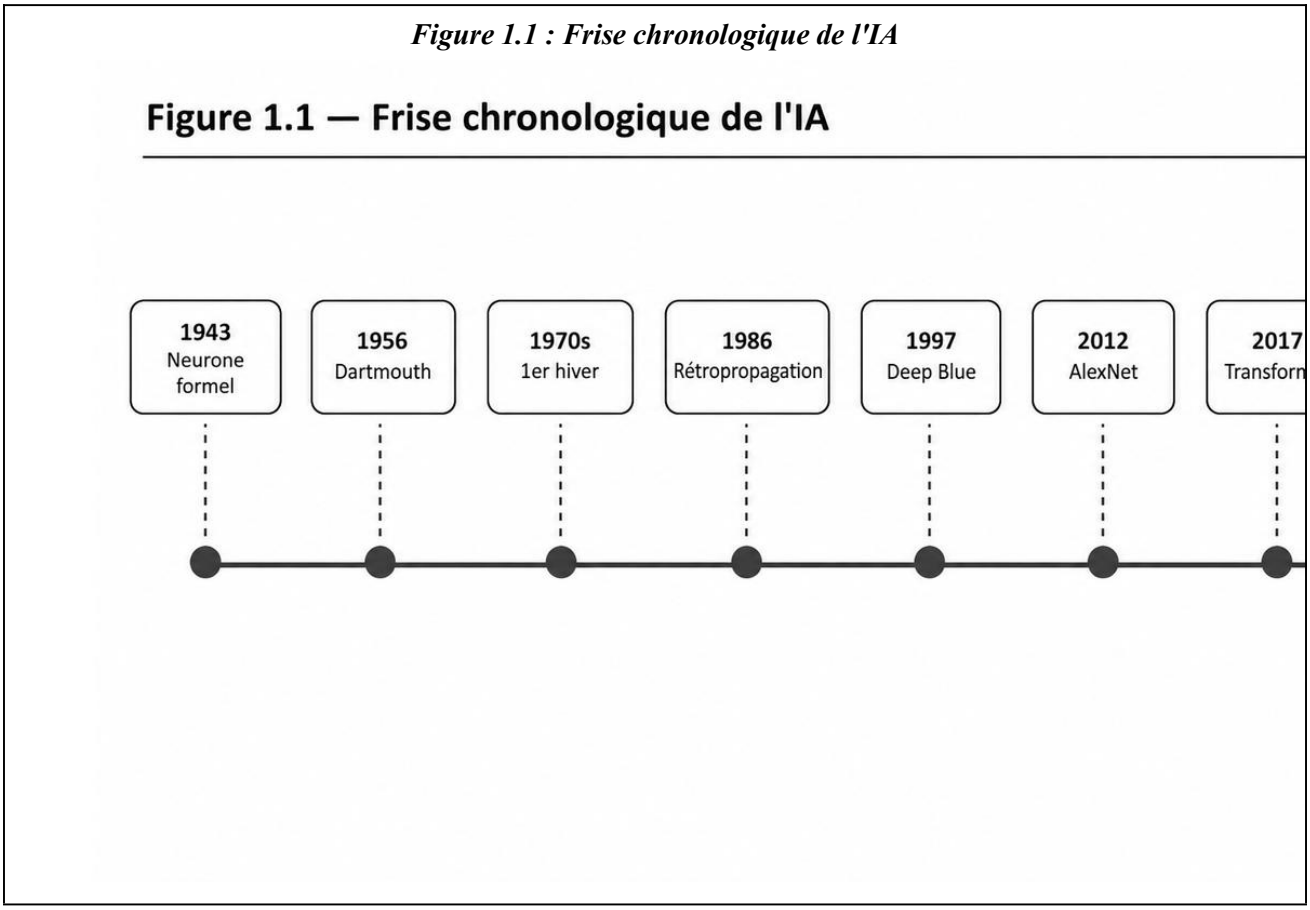
1.3.4 2017 : Les Transformers et la révolution du langage

En 2017, des chercheurs de Google publient un article intitulé « Attention is All You Need » qui propose une nouvelle architecture de réseau de neurones : le Transformer [5]. Cette architecture, spécialement efficace pour les séquences (textes, mais aussi sons, génomes), va révolutionner le traitement du langage. C'est sur les Transformers que reposent ChatGPT, Claude, Gemini, Llama, DeepSeek et tous les grands modèles de langage contemporains.

1.3.5 2022 : L'IA générative entre dans le grand public

Le 30 novembre 2022, OpenAI lance ChatGPT. En cinq jours, le service dépasse un million d'utilisateurs. En deux mois, cent millions [6]. Pour la première fois, des personnes sans formation technique dialoguent naturellement avec une IA capable de rédiger, traduire, coder, résumer, argumenter. L'année 2022 explique en grande partie l'intérêt de ce cours : l'IA n'est plus une affaire de spécialistes.

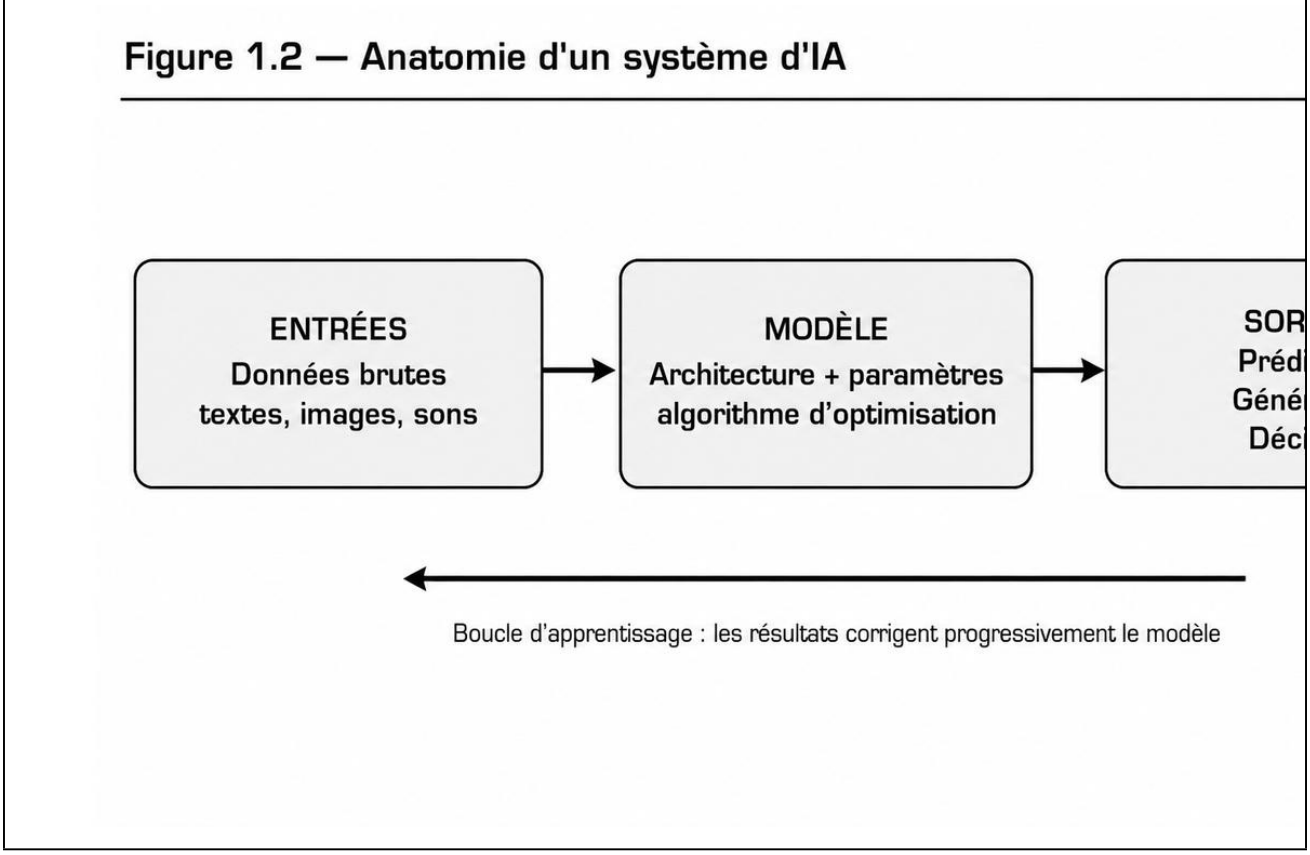
Figure 1.1 : Frise chronologique de l'IA



1.4 Les composantes d'un système d'IA

Tout système d'IA, aussi sophistiqué soit-il, repose sur les mêmes composantes fondamentales. Les comprendre permet de poser les bonnes questions sur n'importe quelle application.

Figure 1.2 : Anatomie d'un système d'IA



1.4.1 Les données : la matière première

Un système d'IA apprend à partir d'exemples. Sans données, il n'y a pas d'IA : exactement comme on ne peut pas devenir médecin sans avoir vu des patients. La qualité, la quantité et la représentativité des données d'entraînement déterminent directement la qualité du modèle produit. Un modèle entraîné uniquement sur des visages d'origine européenne reconnaîtra moins bien les visages d'autres origines. Un modèle entraîné sur des textes sexistes reproduira et amplifiera ces biais. Ce lien direct entre données et résultats est l'une des clés pour comprendre les enjeux politiques de l'IA.

1.4.2 Le modèle : la structure qui apprend

Le modèle est la structure mathématique qui traite les données. Il peut prendre différentes formes selon le problème à résoudre : arbre de décision pour des règles simples, réseau de neurones pour des patterns complexes, modèle de diffusion pour générer des images. L'entraînement consiste à ajuster les paramètres du modèle pour qu'il produise les bonnes sorties sur les données d'entraînement.

1.4.3 L'algorithme d'optimisation : le moteur de l'apprentissage

C'est l'algorithme d'optimisation qui pilote l'apprentissage. Il compare les prédictions du modèle avec les bonnes réponses connues, calcule l'erreur (la « perte »), puis ajuste les paramètres pour réduire cette erreur. Ce processus se répète des millions de fois, sur des milliards d'exemples, jusqu'à ce que le modèle soit suffisamment performant. L'algorithme le plus utilisé s'appelle la descente de gradient stochastique [7].

1.5 Ce que l'IA fait et ce qu'elle ne fait pas

Une erreur fréquente consiste à prêter à l'IA des qualités qu'elle n'a pas : conscience, compréhension, intentions. L'IA contemporaine n'a pas de conscience. Elle ne « sait » pas ce qu'elle fait. Elle calcule. Le fait que ces calculs produisent parfois des résultats spectaculaires n'y change rien.

L'IA peut...	L'IA ne peut pas...
Reconnaître des patterns dans des données massives	Comprendre le sens de ce qu'elle traite
Générer des textes cohérents et convaincants	Avoir des opinions ou des croyances réelles
Battre les humains sur des tâches étroites et définies	Transférer ses capacités d'un domaine à un autre
Traiter des millions de cas simultanément	Faire preuve de sens commun ou de bon sens général
Trouver des corrélations cachées dans les données	Distinguer corrélation et causalité de manière fiable
Optimiser une fonction objectif définie par l'humain	Fixer ses propres objectifs ou valeurs

Tableau 1.2 : Capacités réelles et limites de l'IA contemporaine.

À retenir sur le Chapitre 1

L'IA est une famille de techniques qui permettent à une machine d'apprendre à partir de données, sans que les règles soient explicitement programmées.

La seule forme d'IA qui existe aujourd'hui est l'IA étroite (ANI) : performante sur une tâche, mais incapable de généraliser.

L'IA générale (AGI) et la superintelligence (ASI) sont des objectifs de recherche ou des hypothèses prospectives, pas des réalités.

Tout système d'IA repose sur trois composantes des données, un modèle et un algorithme d'optimisation.

L'IA calcule; elle ne comprend pas, ne ressent pas et n'a pas d'intentions.

1.6 Activités pédagogiques

Activité 1.1 : Identification

Pour chacun des systèmes suivants, indiquez s'il relève de l'IA étroite, de l'IA générale, d'une automatisation classique, ou s'il n'est pas encore réellement déployé. Justifiez en deux phrases.

a) Le filtre anti-spam de votre boîte mail.

b) Un système hypothétique capable de diagnostiquer n'importe quelle maladie ET de rédiger le rapport médical ET de commander les médicaments.

c) Un algorithme de reconnaissance faciale dans un aéroport.

d) Le correcteur orthographique classique qui souligne les fautes.

e) ChatGPT répondant à des questions sur l'histoire algérienne.

Activité 1.2 : Réflexion critique

En janvier 2024, le PDG d'OpenAI, Sam Altman, a déclaré que l'AGI pourrait être créée « dans les prochaines années ». Analysez cette déclaration en identifiant : (a) ce que cela supposerait techniquement, (b) les intérêts économiques et stratégiques qui peuvent motiver une telle annonce, (c) les risques pour la régulation publique si ce type de déclaration est pris au pied de la lettre.

1.7 Résumé du chapitre

Ce premier chapitre a posé les bases conceptuelles indispensables. L'intelligence artificielle désigne un ensemble de techniques permettant à un système de produire des résultats utiles : prédictions, décisions, contenus à partir de données, sans que les règles aient été explicitement programmées. Elle se décline en trois niveaux de capacités hypothétiques, dont seul le premier : l'IA étroite : existe réellement. Son histoire est jalonnée de ruptures techniques (1956, 2012, 2017, 2022) qui reflètent autant des évolutions algorithmiques que des transformations de l'infrastructure matérielle et informationnelle. Comprendre ces bases permet d'aborder le chapitre suivant avec un regard outillé sur le rôle central des données et des algorithmes.

Références bibliographiques : Chapitre 1

Qu'est-ce que l'intelligence artificielle ?

Références citées dans ce chapitre

[1] OECD (2024). Recommendation of the Council on Artificial Intelligence (révision 2024). OECD/LEGAL/0449. Paris : OCDE. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[2] Russell, S. & Norvig, P. (2021). Artificial Intelligence : A Modern Approach, 4e éd. Hoboken : Pearson, 1132 p. ISBN 978-0-13-461099-3.

[3] McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. (2006). « A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence » (1955). AI Magazine, vol. 27, n° 4, p. 12-14. DOI : 10.1609/aimag.v27i4.1904.

[4] Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017). « ImageNet classification with deep convolutional neural networks ». Communications of the ACM, vol. 60, n° 6, p. 84-90. DOI : 10.1145/3065386.

[5] Vaswani, A. et al. (2017). « Attention Is All You Need ». Advances in Neural Information Processing Systems, vol. 30, p. 5998-6008. arXiv : 1706.03762. DOI : 10.48550/arXiv.1706.03762.

[6] Hu, K. (2023). « ChatGPT sets record for fastest-growing user base : analyst note ». Reuters, 2 février 2023. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

[7] Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. Cambridge : MIT Press, 800 p. ISBN 978-0-262-03561-3. Disponible en ligne : https://www.deeplearningbook.org.

Complément du Chapitre 1 : Tour d'horizon des applications actuelles

1.6 L'IA en chiffres : quelques repères

Quelques données de cadrage permettent de situer l'ampleur du phénomène. Le marché mondial de l'IA était estimé à 196 milliards de dollars en 2024 et devrait dépasser 800 milliards d'ici 2030 selon les projections de McKinsey [1]. En 2024, ChatGPT comptait plus de 200 millions d'utilisateurs actifs hebdomadaires. L'entraînement de GPT-4 aurait mobilisé plus de 25 000 processeurs graphiques Nvidia A100 pendant plusieurs mois, pour un coût estimé entre 50 et 100 millions de dollars [2]. Ces chiffres illustrent à eux seuls les enjeux économiques et géopolitiques en jeu.

1.7 Panorama des grandes catégories d'applications

Il peut être utile, avant d'entrer dans les détails techniques des chapitres suivants, d'avoir un panorama des catégories d'applications les plus répandues. Ce panorama permet de situer les usages que vous rencontrerez tout au long du polycopié.

Catégorie	Exemples représentatifs	Chapitres concernés
Traitement du langage	Traduction automatique, chatbots, résumé, analyse de sentiments	5, 6, 10
Vision par ordinateur	Reconnaissance faciale, imagerie médicale, conduite autonome, satellite	6, 9, 11
Systèmes de recommandation	Réseaux sociaux, streaming, e-commerce, fil d'actualité	6, 10
Détection d'anomalies	Fraude bancaire, cybersécurité, contrôle qualité, surveillance épidémique	7, 8, 9
Génération de contenu	Images, textes, musique, vidéos, deepfakes	5, 10
Aide à la décision	Justice prédictive, attribution de crédits, scoring social	7, 9
Robotique et automatisation	Industrie, logistique, agriculture, chirurgie	8
Prévision et modélisation	Météo, épidémies, marchés financiers, résultats électoraux	9, 10, 11

Tableau 1.3 : Panorama des catégories d'applications de l'IA.

1.8 L'IA que vous utilisez déjà sans le savoir

L'intelligence artificielle est présente dans de nombreux outils du quotidien que l'on n'associe pas spontanément à cette technologie. Le correcteur orthographique de votre traitement de texte intègre des modèles de langage simples. Le filtre anti-spam de votre boîte mail est un classifieur entraîné sur des millions d'exemples. La reconnaissance vocale qui transcrit vos messages audio utilise un réseau de neurones récurrent ou un Transformer. La détection de fraude sur votre carte bancaire s'appuie sur un modèle de détection d'anomalies mis à jour en temps réel. Le moteur de recherche qui classe les résultats de vos requêtes applique des dizaines de modèles d'IA successifs. Avant même d'atteindre une application sophistiquée, l'IA structure déjà votre environnement numérique quotidien.

Références bibliographiques : Chapitre 1-Complément

Tour d'horizon des applications actuelles

Références complémentaires

[1] McKinsey Global Institute (2023). The Economic Potential of Generative AI : The Next Productivity Frontier. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai.

[2] Patel, D. & Ahmad, A. (2023). « GPT-4 Architecture, Infrastructure, Training Dataset, Costs, Vision, MoE ». SemiAnalysis. https://www.semianalysis.com/p/gpt-4-architecture-infrastructure.

Chapitre 2 : Données, algorithmes et apprentissage

Objectifs pédagogiques

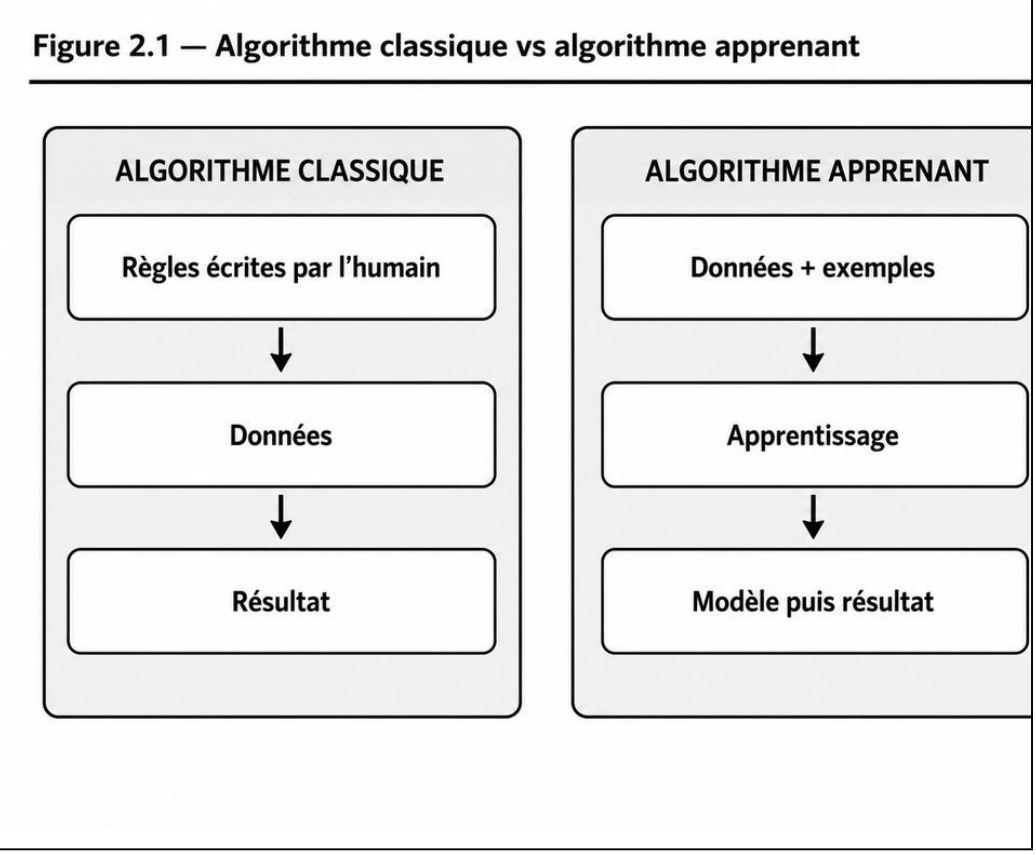
- Comprendre ce qu'est réellement un algorithme et le distinguer d'une règle explicite.
- Saisir le rôle central des données dans tout système d'IA.
- Identifier les trois grandes familles d'apprentissage automatique.
- Comprendre la notion de Big Data et ses cinq dimensions.
- Lire un schéma pipeline de traitement de données sans formation technique.

2.1 L'algorithme : l'idée la plus simple mal comprise

Un algorithme, c'est une suite d'instructions précises, dans un ordre défini, pour résoudre un problème. Rien de mystérieux. La recette de la chorba est un algorithme. Le plan d'évacuation d'un bâtiment en est un autre. Ce qui différencie un algorithme informatique, c'est sa vitesse d'exécution : un ordinateur peut répéter des milliards d'opérations par seconde, sans fatigue, sans erreur d'inattention.

La distinction décisive pour comprendre l'IA moderne oppose deux types d'algorithmes. Le premier type, classique, suit des règles écrites par un humain : « si la température dépasse 38°C, déclencher l'alarme ». Prévisible, transparent, mais limité aux situations anticipées. Le second type, fondé sur l'apprentissage, déduit ses propres règles à partir d'exemples : on lui montre des milliers de cas « alarme déclenchée » et « alarme non déclenchée », et il apprend à reconnaître quand déclencher l'alarme lui-même, parfois en détectant des patterns que l'humain n'aurait pas formulés.

Figure 2.1 : Algorithme classique vs algorithme apprenant



2.2 Les données : sans elles, pas d'IA

Les données sont à l'IA ce que la farine est au pain : on ne peut rien faire sans elles, et la qualité du résultat dépend directement de leur qualité. Un modèle d'IA apprend à partir d'exemples. Ces exemples constituent ce qu'on appelle le jeu de données d'entraînement. Plus il est grand, varié et représentatif, meilleur sera le modèle : en général.

2.2.1 Les types de données

Les systèmes d'IA peuvent traiter des données très diverses. Les données structurées sont organisées en tableaux avec des colonnes définies : feuilles Excel, bases de données SQL. Les données non structurées représentent 80 à 90 % des données mondiales [1] : textes, images, vidéos, sons, publications sur les réseaux sociaux. Les modèles les plus récents sont capables de travailler simultanément avec plusieurs types de données : on parle alors de modèles multimodaux.

2.2.2 Le Big Data et ses cinq dimensions

Le terme « Big Data » ne désigne pas simplement de grandes quantités de données. Il décrit un ensemble de caractéristiques qui rendent le traitement traditionnel inadapté [2]. La littérature les résume en cinq « V » :

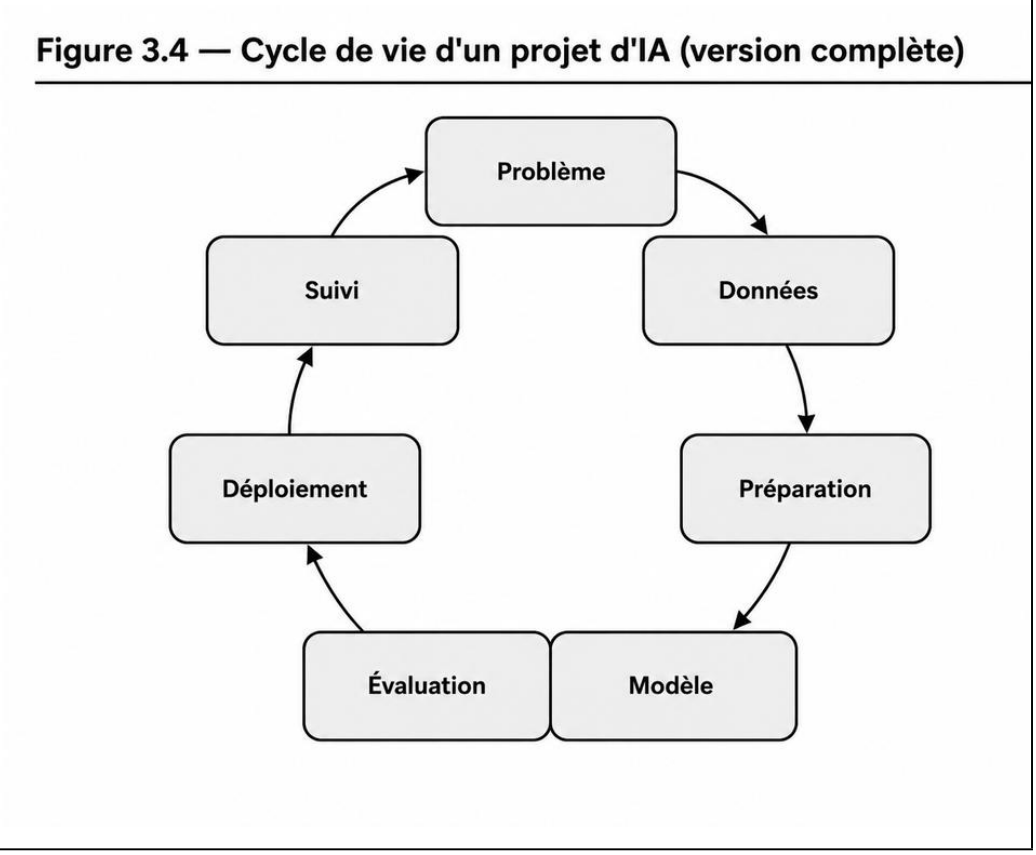
Dimension	Définition	Exemple concret en politique
Volume	Quantité de données générées	Millions de tweets lors d'une élection
Vélocité	Vitesse de génération et traitement	Flux en temps réel des réseaux sociaux
Variété	Diversité des formats	Textes, images, vidéos, géolocalisation
Véracité	Fiabilité et exactitude des données	Présence de fausses informations dans les corpus
Valeur	Utilité exploitable des données	Profils psychologiques pour le ciblage électoral

Tableau 2.1 : Les cinq dimensions du Big Data avec exemples politiques. Source : adapté de [2].

2.2.3 Le pipeline de traitement des données

Avant d'alimenter un modèle d'IA, les données passent par plusieurs étapes de préparation. Ce processus, appelé pipeline, est souvent la partie la plus longue et la plus coûteuse d'un projet d'IA : on estime qu'il représente entre 60 et 80 % du travail total [3].

Figure 2.2 : Pipeline de traitement des données



2.3 Les trois familles d'apprentissage automatique

L'apprentissage automatique (Machine Learning) regroupe un ensemble de méthodes permettant à un système d'apprendre à partir de données. Trois grandes familles se distinguent selon le type d'information disponible lors de l'entraînement [4].

2.3.1 L'apprentissage supervisé

C'est la famille la plus répandue. Elle suppose que chaque exemple du jeu d'entraînement est accompagné d'une étiquette correcte : la bonne réponse. On montre au modèle des milliers de paires (entrée → réponse attendue) et il apprend à associer les entrées aux bonnes réponses.

Exemple : Détection de spam

Un service de messagerie possède un historique de 500 000 courriels, chacun marqué « spam » ou « légitime » par des humains. On entraîne un modèle sur ces exemples. Après entraînement, le modèle peut classer automatiquement les nouveaux courriels : même ceux formulés différemment des exemples vus. C'est de l'apprentissage supervisé : la supervision vient des étiquettes humaines.

Applications courantes : détection de spam, reconnaissance vocale, traduction automatique, diagnostic médical assisté, prévision de résultats électoraux.

2.3.2 L'apprentissage non supervisé

Cette famille travaille sur des données sans étiquettes. Le système doit découvrir lui-même des structures, des groupes ou des patterns cachés. L'humain ne dit pas « voici la bonne réponse » : il attend que la machine révèle une organisation qu'il n'avait pas anticipée.

Exemple : Segmentation électorale

Un parti politique dispose de données démographiques et comportementales sur des millions d'électeurs, sans étiquettes prédéfinies. Un algorithme de clustering (regroupement) identifie automatiquement cinq profils distincts d'électeurs, chacun avec ses caractéristiques propres. Le parti peut alors adapter ses messages à chaque profil. Aucune étiquette n'était fournie au départ : la machine a découvert la structure.

2.3.3 L'apprentissage par renforcement

Cette famille fonctionne sur le principe de l'essai-erreur, avec récompenses et pénalités. L'agent (le système d'IA) effectue des actions dans un environnement, observe les conséquences, et ajuste sa stratégie pour maximiser ses récompenses à long terme.

Exemple : AlphaGo

Pour apprendre à jouer au Go, AlphaGo a joué des millions de parties contre lui-même. Chaque victoire était une récompense, chaque défaite une pénalité. Sans qu'aucune règle stratégique n'ait été codée explicitement, le système a découvert des stratégies que les meilleurs joueurs humains n'avaient pas envisagées [5].

Famille	Données requises	Objectif	Exemples d'applications
Supervisé	Données étiquetées	Prédire une étiquette pour de nouvelles données	Spam, diagnostic, traduction
Non supervisé	Données non étiquetées	Découvrir des structures cachées	Segmentation, détection d'anomalies
Par renforcement	Environnement + système de récompenses	Optimiser une stratégie par essai-erreur	Jeux, robotique, optimisation

Tableau 2.2 : Comparaison des trois familles d'apprentissage automatique. Source : adapté de [4].

<u>À retenir sur le Chapitre 2</u>

Un algorithme est une suite d'instructions précises. L'IA moderne utilise des algorithmes qui apprennent leurs propres règles à partir de données.

Les données sont la matière première de l'IA. Leur qualité, quantité et représentativité déterminent directement la performance et l'équité du modèle.

Le Big Data se caractérise par cinq dimensions : Volume, Vélocité, Variété, Véracité, Valeur.

Trois familles d'apprentissage existent : supervisé (avec étiquettes), non supervisé (sans étiquettes), par renforcement (récompenses et pénalités).

La préparation des données représente 60 à 80 % du travail dans un projet d'IA.

Activité 2.1 : Classification

Pour chacun des scénarios suivants, identifiez la famille d'apprentissage utilisée et justifiez :

a) Un système qui analyse des milliers de discours politiques étiquetés 'populiste' ou 'non populiste' pour en détecter automatiquement les caractéristiques.

b) Un algorithme qui analyse les comportements de vote de milliers de citoyens pour identifier spontanément des groupes aux profils similaires.

c) Un drone militaire qui apprend à éviter les obstacles en simulant des milliers de trajectoires et en recevant des points chaque fois qu'il évite un obstacle.

Références bibliographiques : Chapitre 2

Données, algorithmes et apprentissage

Références citées

[1] IDC (2023). Data Creation and Replication Will Grow at a Faster Rate than Installed Storage Capacity. Framingham : IDC. https://www.idc.com/getdoc.jsp?containerId=prUS50396223.

[2] Laney, D. (2001). 3D Data Management : Controlling Data Volume, Velocity and Variety. Application Delivery Strategies, META Group. Note de recherche. https://www.gartner.com (archivé).

[3] Anaconda (2022). The State of Data Science 2022. https://www.anaconda.com/state-of-data-science-report.

[4] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. New York : Springer, 738 p. ISBN 978-0-387-31073-2.

[5] Silver, D. et al. (2016). « Mastering the game of Go with deep neural networks and tree search ». Nature, vol. 529, n° 7587, p. 484-489. DOI : 10.1038/nature16961.

Complément du Chapitre 2 : Qualité des données et étiquetage

2.4 La qualité des données : un enjeu souvent sous-estimé

Un principe bien connu en informatique s'énonce ainsi : « Garbage in, garbage out » : ordures en entrée, ordures en sortie. En Machine Learning, ce principe prend une dimension politique : les données de mauvaise qualité produisent des modèles biaisés ou inexacts, et ces modèles prennent ou orientent des décisions qui affectent des personnes réelles.

2.4.1 Les dimensions de la qualité des données

Dimension	Définition	Conséquence si négligée
Complétude	Toutes les valeurs importantes sont présentes	Le modèle apprend des patterns tronqués
Exactitude	Les données reflètent la réalité	Apprentissage de fausses régularités
Cohérence	Les données sont uniformes dans leur format	Erreurs de traitement, bruit dans le modèle
Fraîcheur	Les données sont récentes et à jour	Modèle incapable de gérer les évolutions récentes
Représentativité	Tous les groupes pertinents sont présents	Biais contre les groupes sous-représentés
Pertinence	Les données sont utiles pour la tâche visée	Modèle apprenant des corrélations parasites

Tableau 2.3 : Les six dimensions de la qualité des données en IA.

2.4.2 L'étiquetage : le travail humain invisible

L'apprentissage supervisé repose sur des données étiquetées : c'est-à-dire que chaque exemple est associé à la bonne réponse par un humain. Ce travail d'annotation est massif, répétitif et souvent externalisé à des travailleurs précaires dans les pays en développement. Une étude du magazine Time de 2023 a révélé que des annotateurs kenyans, ougandais et tanzaniens travaillaient pour OpenAI à moins de 2 dollars de l'heure pour identifier des contenus toxiques afin d'entraîner les filtres de sécurité de ChatGPT [1].

Ce travail invisible conditionne la qualité et les biais des modèles. Si les annotateurs ont eux-mêmes des biais culturels ou si les instructions d'annotation sont ambiguës, ces biais se retrouvent dans le modèle. La chaîne de production d'une IA performante implique donc bien plus que des ingénieurs et des serveurs : elle mobilise une main-d'œuvre mondiale aux conditions souvent précaires.

Analogie : L'entrepôt de données

Imaginez un supermarché qui veut créer un système de recommandation de recettes basé sur les achats. Si 90% de sa clientèle vit dans une grande ville et achète des produits importés, le modèle sera excellent pour recommander des recettes aux citadins aisés, mais inutile pour les clients ruraux qui achètent des produits locaux peu représentés dans les données. L'algorithme n'est pas biaisé parce qu'il est 'méchant' : il est biaisé parce que les données qui l'ont nourri étaient biaisées.

2.5 Jeux de données publics de référence

La communauté scientifique a constitué des jeux de données publics qui servent de benchmarks pour comparer les performances des modèles. Les connaître permet de comprendre sur quelles bases les systèmes d'IA sont évalués et de saisir pourquoi ces benchmarks ont aussi leurs limites.

Jeu de données	Domaine	Taille	Utilisation
ImageNet	Vision : classification d'images	14 millions d'images, 20 000 catégories	Benchmark mondial depuis 2010 : AlexNet en 2012
MNIST	Vision : chiffres manuscrits	70 000 images de chiffres 0-9	Premier benchmark d'apprentissage : encore utilisé
Common Crawl	NLP : pages web	Plusieurs pétaoctets de texte	Données d'entraînement des grands LLM
SQuAD	NLP : questions-réponses	100 000 questions sur Wikipedia	Évaluation de la compréhension du langage
GLUE / SuperGLUE	NLP : plusieurs tâches	Ensemble de 9 tâches linguistiques	Benchmark global des modèles de langage
OpenAssistant	NLP : conversations	161 000 échanges en 35 langues	Entraînement de chatbots multilingues

Tableau 2.4 : Principaux jeux de données de référence en IA. Sources : sites officiels des datasets.

Références bibliographiques : Chapitre 2-Complément

Qualité des données et étiquetage

Références complémentaires

[1] Perrigo, B. (2023). « Exclusive : OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic ». Time Magazine, 18 janvier 2023. https://time.com/6247678/openai-chatgpt-kenya-workers/.

Extension Chapitre 2 : Les algorithmes classiques en détail

2.6 Zoom sur cinq algorithmes fondamentaux

Après avoir vu les familles d'apprentissage, approfondissons cinq algorithmes concrets que vous rencontrerez souvent dans les applications politiques et administratives. Pour chacun, l'explication reste intuitive : aucune formule n'est requise, mais suffisamment précise pour que vous puissiez en analyser les usages réels.

2.6.1 La régression logistique

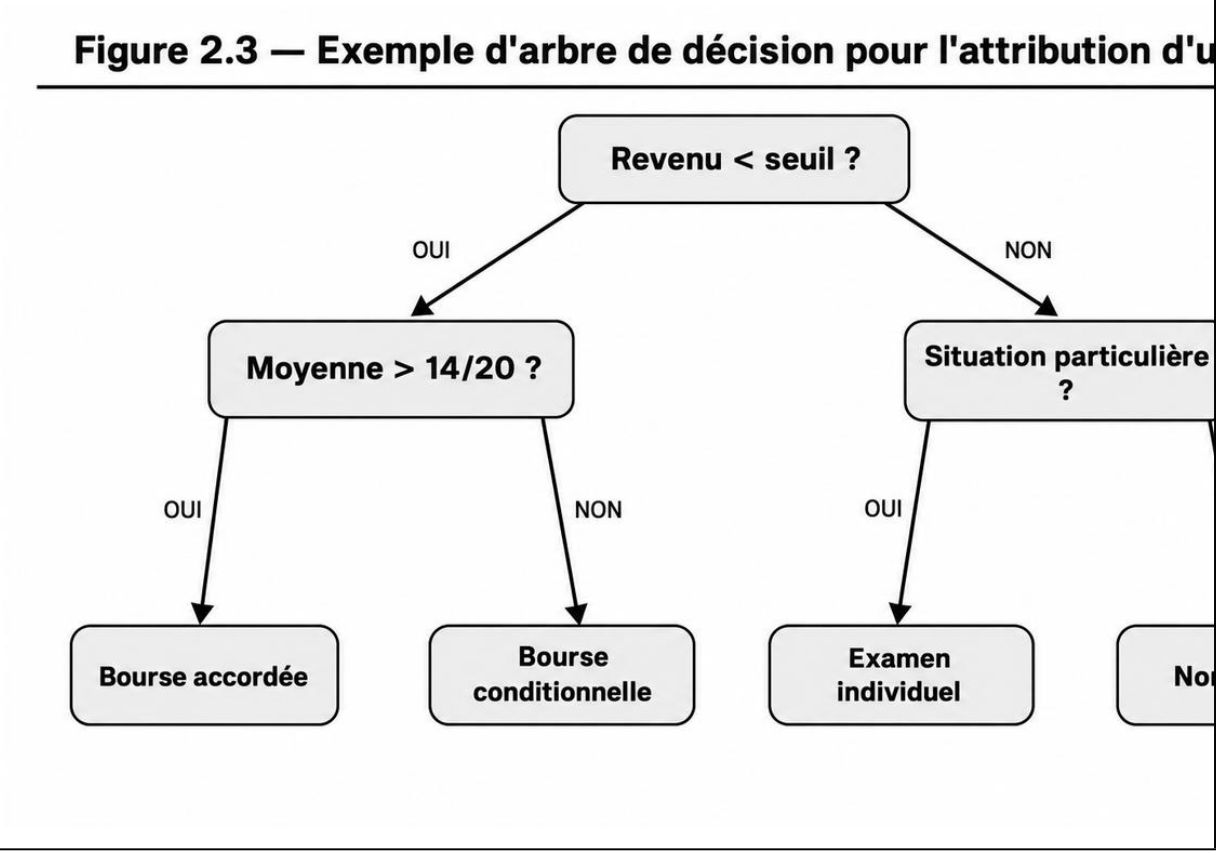
Malgré son nom qui contient le mot « régression », la régression logistique est un algorithme de classification. Elle calcule la probabilité qu'un exemple appartienne à une catégorie donnée. Pour un filtre anti-spam, elle calcule : « quelle est la probabilité que ce courriel soit un spam ? ». Si cette probabilité dépasse un seuil (par exemple 0,7), le courriel est classé comme spam.

Son atout majeur est l'interprétabilité : on peut directement voir quel mot contribue le plus à la décision. C'est pourquoi elle reste populaire dans des contextes où la justification des décisions est requise : accès aux prestations sociales, évaluation de risques médicaux. Sa limite est sa linéarité : elle ne capture pas bien des relations complexes entre variables.

2.6.2 L'arbre de décision

Un arbre de décision pose une série de questions binaires sur les données, en suivant les branches selon les réponses, jusqu'à atteindre une décision finale. Prenons l'exemple d'un système d'attribution de bourses : « Le revenu du foyer est-il inférieur à X ? Si oui, le niveau académique est-il supérieur à Y ? Si oui, attribuer la bourse. »

Figure 2.3 : Exemple d'arbre de décision pour l'attribution d'une bourse



2.6.3 La forêt aléatoire

La forêt aléatoire (Random Forest) est un ensemble de dizaines ou de centaines d'arbres de décision, chacun entraîné sur un sous-ensemble aléatoire des données et des variables. La décision finale résulte du vote majoritaire de tous les arbres. Ce mécanisme réduit considérablement le surapprentissage et améliore la robustesse. C'est l'algorithme préféré pour la détection de fraude fiscale, le scoring de crédit et le profilage de risques dans les administrations.

2.6.4 Les k plus proches voisins (k-NN)

Le k-NN est l'algorithme le plus intuitif qui soit : pour classifier un nouvel exemple, on cherche ses k voisins les plus proches dans les données d'entraînement, et on lui attribue la catégorie majoritaire parmi ces voisins. Analogie directe : si vous déménagez dans un nouveau quartier et que 7 de vos 10 voisins les plus proches sont fonctionnaires, l'algorithme prédit que vous l'êtes aussi. Sa limite : il est lent sur de grandes bases de données et sensible aux variables non pertinentes.

2.6.5 Le k-Means (regroupement)

Le k-Means est l'algorithme de regroupement non supervisé le plus utilisé. On lui demande de diviser les données en k groupes, en minimisant la distance intra-groupe. Exemple d'usage politique : donner à l'algorithme les données de comportement en ligne de millions d'électeurs et lui demander de trouver 5 groupes naturels : il identifiera lui-même des profils cohérents sans qu'on les ait définis à l'avance. C'est exactement ce que fait le micro-ciblage électoral à grande échelle.

2.7 Les bases de données vectorielles : l'infrastructure de la RAG

Une notion en plein essor mérite d'être introduite ici : les bases de données vectorielles. Elles stockent des données (textes, images) sous forme de vecteurs numériques (embeddings) et permettent de retrouver rapidement les éléments les plus similaires à une requête. Elles sont au cœur des architectures RAG (Retrieval-Augmented Generation) mentionnées au chapitre 6 : quand un chatbot « recherche dans une base de connaissances », il interroge une base de données vectorielle pour trouver les passages les plus pertinents, puis les passe au LLM pour générer la réponse.

Pour une administration algérienne, une base de données vectorielle permet de construire un chatbot capable de répondre aux questions des usagers en se basant uniquement sur les textes juridiques officiels sans halluciner des lois qui n'existent pas.

Chapitre 3 : Machine Learning et Deep Learning

Objectifs pédagogiques

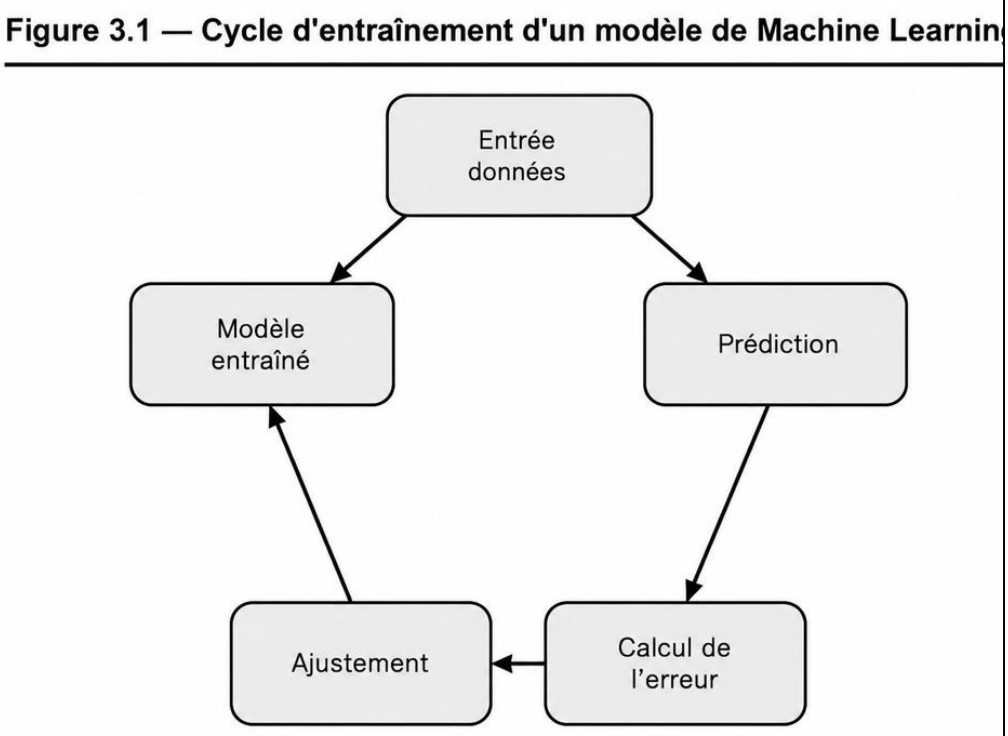
- Comprendre le principe fondamental du Machine Learning : apprendre par l'erreur.
- Saisir ce que signifie 'entraîner un modèle' sans formules mathématiques.
- Distinguer Machine Learning et Deep Learning.
- Comprendre la notion de surapprentissage (overfitting) et ses conséquences.
- Identifier les principaux algorithmes de ML et leurs usages politiques.

3.1 Machine Learning : apprendre de ses erreurs

Imaginez que vous apprenez à distinguer une photo d'un dattier d'une photo d'un olivier. Au début, vous faites des erreurs. Quelqu'un vous corrige. Progressivement, vous intériorisez les caractéristiques qui distinguent les deux arbres. Vous ne mémorisez pas les photos vues : vous généralisez. C'est exactement ce que fait le Machine Learning, mais en numérique et à une échelle bien plus grande.

Un modèle de Machine Learning commence avec des paramètres aléatoires. Il fait une prédiction, compare avec la bonne réponse, mesure l'erreur, et ajuste ses paramètres pour réduire cette erreur. On répète ce processus des millions de fois sur des millions d'exemples. À la fin, le modèle produit des prédictions correctes sur des données qu'il n'a jamais vues : s'il a bien généralisé.

Figure 3.1 : Cycle d'entraînement d'un modèle de Machine Learning



3.1.1 Qu'est-ce qu'un paramètre ?

Les paramètres sont les « boutons de réglage » du modèle. Lors de l'entraînement, l'algorithme tourne ces boutons pour minimiser les erreurs. Un modèle de régression linéaire simple n'a que deux paramètres. Un réseau de neurones modeste en a des milliers. GPT-4, le grand modèle de langage d'OpenAI, en aurait plus de mille milliards. Le nombre de paramètres détermine la capacité d'apprentissage, mais aussi la consommation de ressources et les risques de surapprentissage.

3.1.2 Surapprentissage : quand le modèle apprend trop bien

Un des problèmes classiques du Machine Learning s'appelle le surapprentissage (overfitting en anglais). Il survient quand le modèle apprend les données d'entraînement par cœur, y compris leurs bizarreries et leurs erreurs au lieu de dégager des règles générales. Un modèle en surapprentissage performe parfaitement sur ses données d'entraînement, mais échoue sur des données nouvelles.

Analogie : L'étudiant qui apprend par cœur

Un étudiant qui mémorise mot pour mot les sujets des années précédentes sans comprendre les notions sous-jacentes sera en difficulté face à un sujet légèrement différent. C'est le surapprentissage : excellente performance sur le connu, échec sur l'inédit. La solution, pour l'IA comme pour l'étudiant, est de comprendre les principes plutôt que de mémoriser les exemples.

3.2 Les principaux algorithmes de Machine Learning

Le Machine Learning rassemble une famille d'algorithmes aux approches variées. Voici les plus importants, avec leurs applications concrètes dans le domaine politique et administratif.

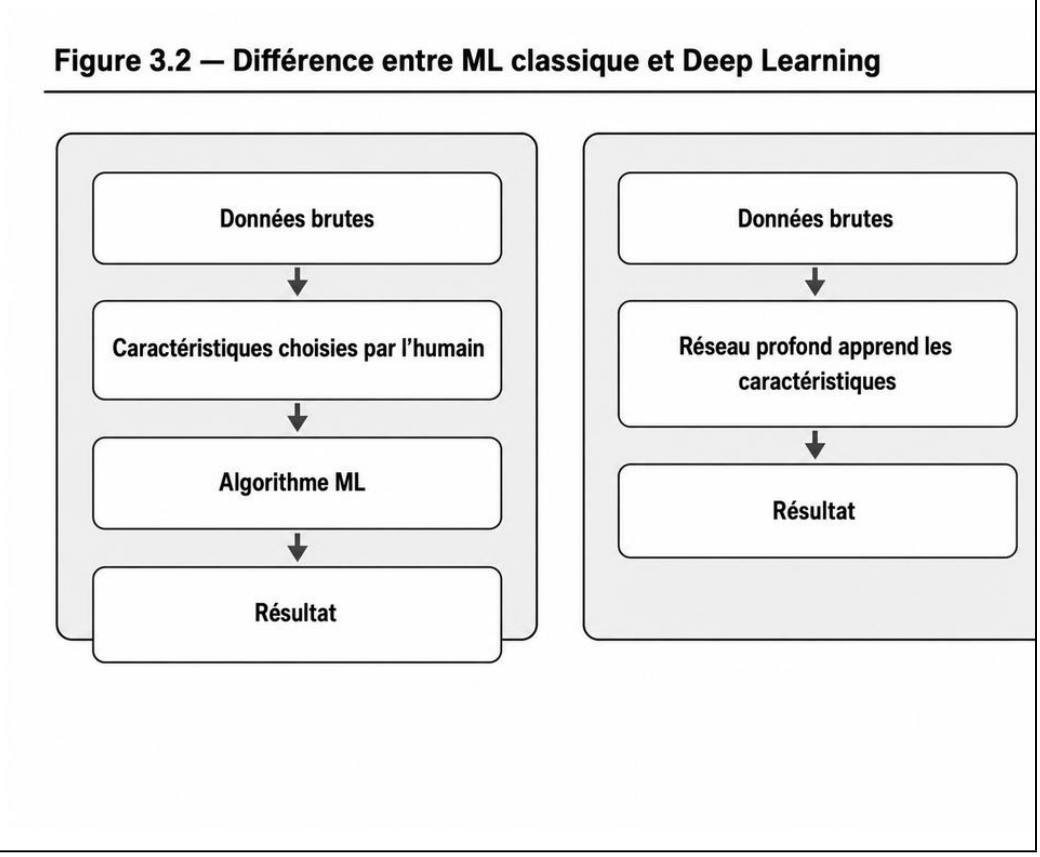
Algorithme	Principe simplifié	Application politique / administrative
Régression logistique	Calcule une probabilité d'appartenance à une catégorie	Prédire le résultat d'un vote (pour/contre une loi)
Arbre de décision	Série de questions binaires imbriquées	Évaluer l'éligibilité à une prestation sociale
Forêt aléatoire	Ensemble d'arbres de décision indépendants	Détection de fraude fiscale, scoring de risque
SVM	Trouve la frontière optimale entre catégories	Classification de documents administratifs
Réseau de neurones	Couches de neurones artificiels interconnectés	Reconnaissance vocale, analyse d'images satellites
K-Means	Regroupe les données en K clusters	Segmentation de l'électorat, profilage d'usagers

Tableau 3.1 : Principaux algorithmes de ML et leurs usages. Source : adapté de [1].

3.3 Deep Learning : les réseaux profonds

Le Deep Learning (apprentissage profond) est une sous-famille du Machine Learning qui utilise des réseaux de neurones à plusieurs couches successives : d'où l'adjectif « profond ». Sa particularité est de pouvoir apprendre des représentations hiérarchiques des données sans que l'humain ait à extraire manuellement les caractéristiques pertinentes.

Figure 3.2 : Différence entre ML classique et Deep Learning



3.3.1 Pourquoi le Deep Learning fonctionne

Un réseau de neurones profond apprend des représentations hiérarchiques. Prenons l'exemple de la reconnaissance d'un visage. La première couche du réseau apprend à détecter des contours simples (lignes horizontales, verticales). La deuxième couche combine ces contours pour détecter des formes (yeux, nez, bouche). La troisième couche combine ces formes pour reconnaître un visage entier. Chaque couche construit sur la précédente, permettant une abstraction progressive : exactement comme le cerveau visuel humain, dont cette architecture s'inspire [2].

3.3.2 L'empreinte énergétique du Deep Learning

Le Deep Learning a un coût : l'entraînement d'un grand modèle consomme une quantité astronomique d'énergie. Une étude de 2019 a estimé que l'entraînement d'un modèle de traitement du langage de grande taille émettait autant de CO_2 que cinq voitures pendant toute leur durée de vie [3]. Ce coût environnemental est un enjeu politique réel, qui soulève des questions sur qui a accès à ces technologies et qui en supporte le coût.

À retenir sur le Chapitre 3

Le Machine Learning apprend par l'erreur : le modèle ajuste ses paramètres pour minimiser son erreur de prédiction.

Le surapprentissage (overfitting) survient quand le modèle mémorise les données au lieu de généraliser.

Le Deep Learning utilise des réseaux de neurones profonds qui apprennent automatiquement les caractéristiques pertinentes.

La différence avec le ML classique : en DL, l'humain ne choisit pas quoi regarder : le réseau le découvre.

L'entraînement de grands modèles de Deep Learning a un coût énergétique et financier considérable.

Activité 3.1 : Analyse de cas

En 2018, Amazon a abandonné un système d'IA de sélection de CV car il discriminait les candidatures féminines. En utilisant les notions de ce chapitre, expliquez : (a) comment un algorithme de ML peut reproduire des discriminations présentes dans les données d'entraînement, (b) pourquoi le surapprentissage peut aggraver ce problème, (c) quelles mesures techniques et organisationnelles auraient pu prévenir cette dérive.

Références bibliographiques : Chapitre 3

Machine Learning et Deep Learning

Références citées

[1] Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3e éd. Sebastopol : O'Reilly Media, 851 p. ISBN 978-1-098-12597-4.

[2] LeCun, Y., Bengio, Y. & Hinton, G. (2015). « Deep learning ». Nature, vol. 521, n° 7553, p. 436-444. DOI : 10.1038/nature14539.

[3] Strubell, E., Ganesh, A. & McCallum, A. (2019). « Energy and Policy Considerations for Deep Learning in NLP ». Proceedings of the 57th Annual Meeting of the ACL, p. 3645-3650. DOI : 10.18653/v1/P19-1355.

Complément du Chapitre 3 : Évaluation et métriques des modèles

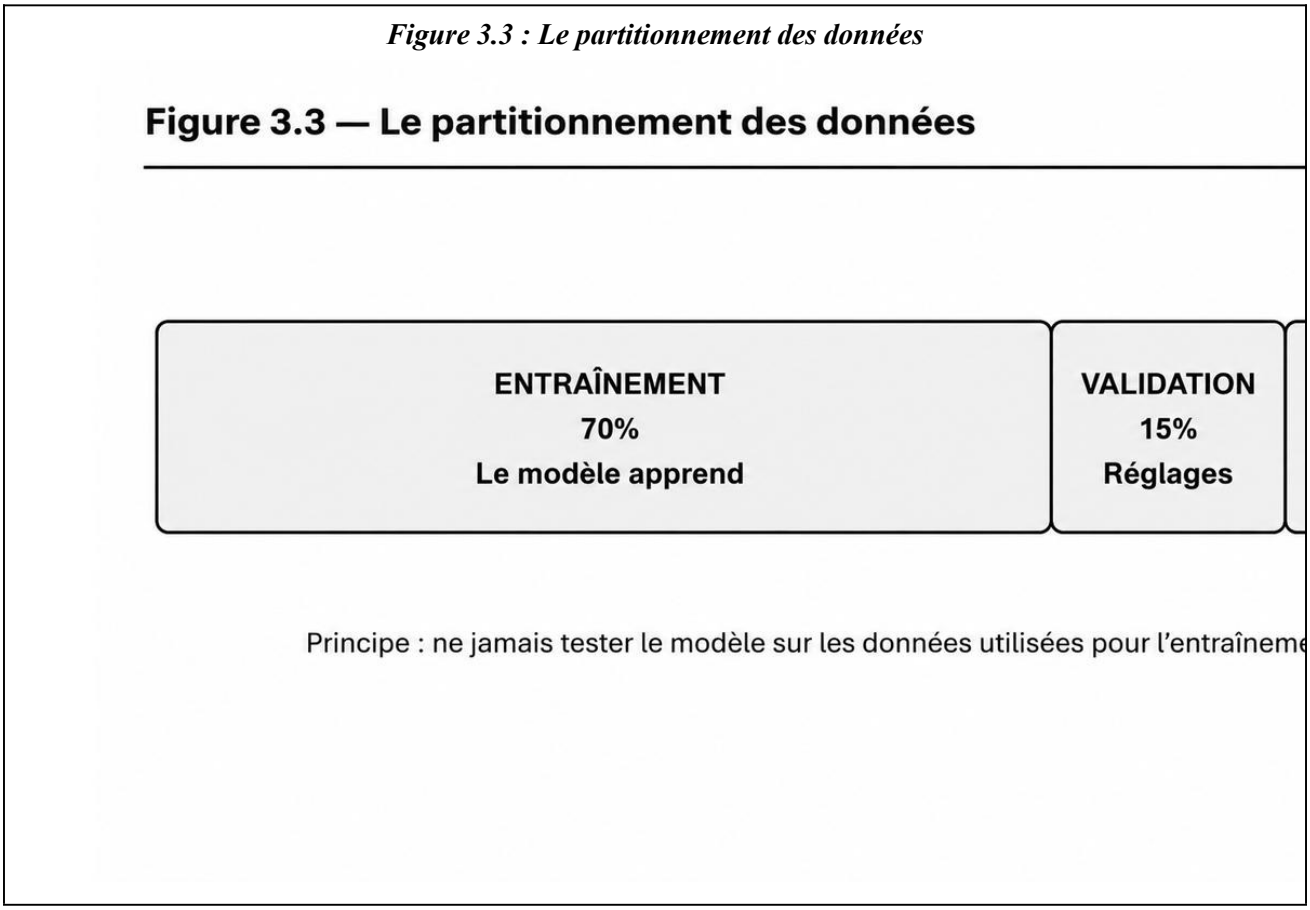
3.4 Comment évaluer la performance d'un modèle ?

Entraîner un modèle n'est que la moitié du travail. L'autre moitié consiste à mesurer sa performance de manière rigoureuse et à distinguer la performance sur les données d'entraînement de la performance sur des données nouvelles. C'est cette seconde mesure qui compte vraiment.

3.4.1 Le partitionnement des données

La pratique standard consiste à diviser les données disponibles en trois ensembles distincts avant tout entraînement. L'ensemble d'entraînement, qui représente environ 70% des données, sert à ajuster les paramètres du modèle. L'ensemble de validation, environ 15%, sert à surveiller la performance pendant l'entraînement et à ajuster les hyperparamètres (la structure du modèle elle-même). L'ensemble de test, environ 15%, n'est utilisé qu'une seule fois, à la toute fin, pour mesurer la performance réelle sur des données jamais vues. Cette règle de séparation stricte est fondamentale : tester un modèle sur ses données d'entraînement revient à faire passer un examen avec les réponses.

Figure 3.3 : Le partitionnement des données



3.4.2 Les métriques d'évaluation

Pour un problème de classification binaire (spam/légitime, malade/sain, fraudeur/honnête), plusieurs métriques s'appliquent selon ce qu'on cherche à optimiser. Comprendre ces métriques est important car le choix de l'une ou de l'autre a des conséquences politiques directes : notamment pour les applications de justice, de médecine ou de crédit.

Métrique	Ce qu'elle mesure	Quand la privilégier
Exactitude (Accuracy)	Proportion de prédictions correctes parmi toutes	Quand les classes sont équilibrées
Précision (Precision)	Parmi les cas classés positifs, combien le sont vraiment	Quand les faux positifs coûtent cher (ex. fausses arrestations)
Rappel (Recall)	Parmi les vrais positifs, combien le modèle en détecte	Quand les faux négatifs coûtent cher (ex. cancer non détecté)
F1-Score	Moyenne harmonique de la précision et du rappel	Compromis équilibré entre précision et rappel
AUC-ROC	Performance globale sur tous les seuils de décision	Comparaison robuste de modèles différents

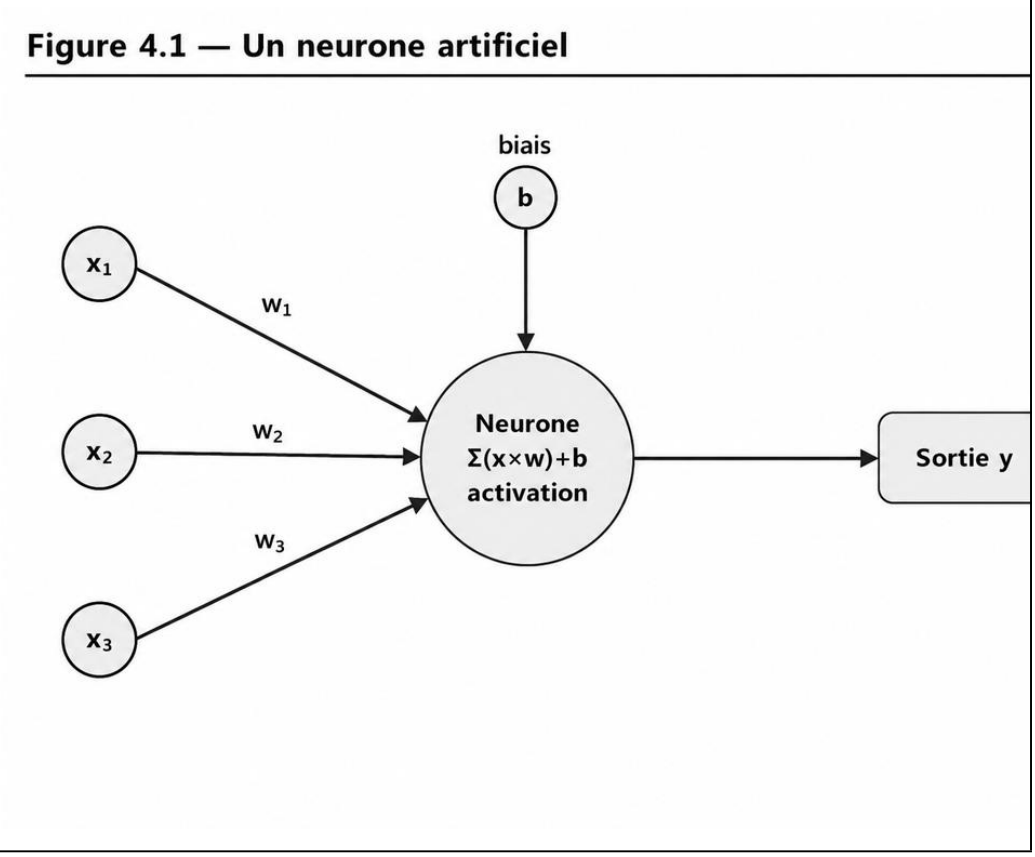
Tableau 3.2 : Métriques d'évaluation pour les problèmes de classification.

Exemple politique : Le coût des erreurs asymétriques

Un système de détection de fraude à l'examen peut commettre deux types d'erreurs : classer un étudiant honnête comme fraudeur (faux positif) ou ne pas détecter un vrai fraudeur (faux négatif). Ces deux erreurs n'ont pas le même coût moral et juridique. Exclure un innocent est une injustice grave; rater un fraudeur est moins grave mais fragilise l'équité. Selon ce qu'on veut prioritairement éviter, on choisit d'optimiser la précision (moins de faux positifs) ou le rappel (moins de faux négatifs). Ce choix est une décision politique, pas seulement technique.

3.5 Le cycle de vie complet d'un projet d'IA

Figure 3.4 : Cycle de vie d'un projet d'IA (version complète)



Chapitre 4 : Réseaux de neurones et modèles de fondation

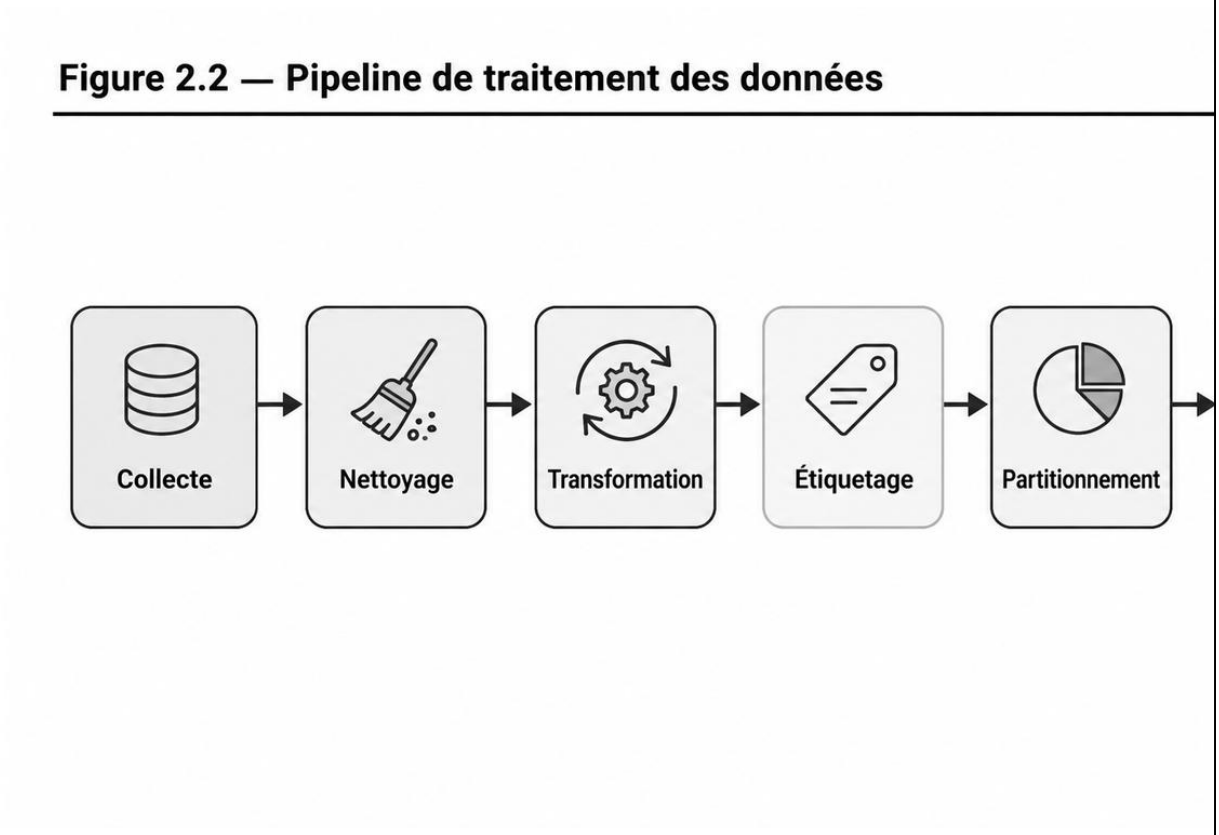
Objectifs pédagogiques

- Comprendre le neurone artificiel sans formule mathématique.
- Saisir comment un réseau de neurones apprend par rétropropagation.
- Distinguer les architectures majeures : CNN, RNN, Transformer.
- Comprendre ce qu'est un modèle de fondation et pourquoi GPT-4 consomme autant de ressources.
- Identifier les grandes familles de modèles disponibles en 2025-2026.

4.1 Le neurone artificiel : l'unité de base

Tout réseau de neurones est construit à partir d'une brique élémentaire : le neurone artificiel. Son fonctionnement est inspiré, de manière très approximative, du neurone biologique. Un neurone artificiel reçoit plusieurs signaux en entrée, les pondère (attribue un poids à chacun), les additionne, et décide s'il « s'active » en passant ou non un signal à la couche suivante.

Figure 4.1 : Un neurone artificiel



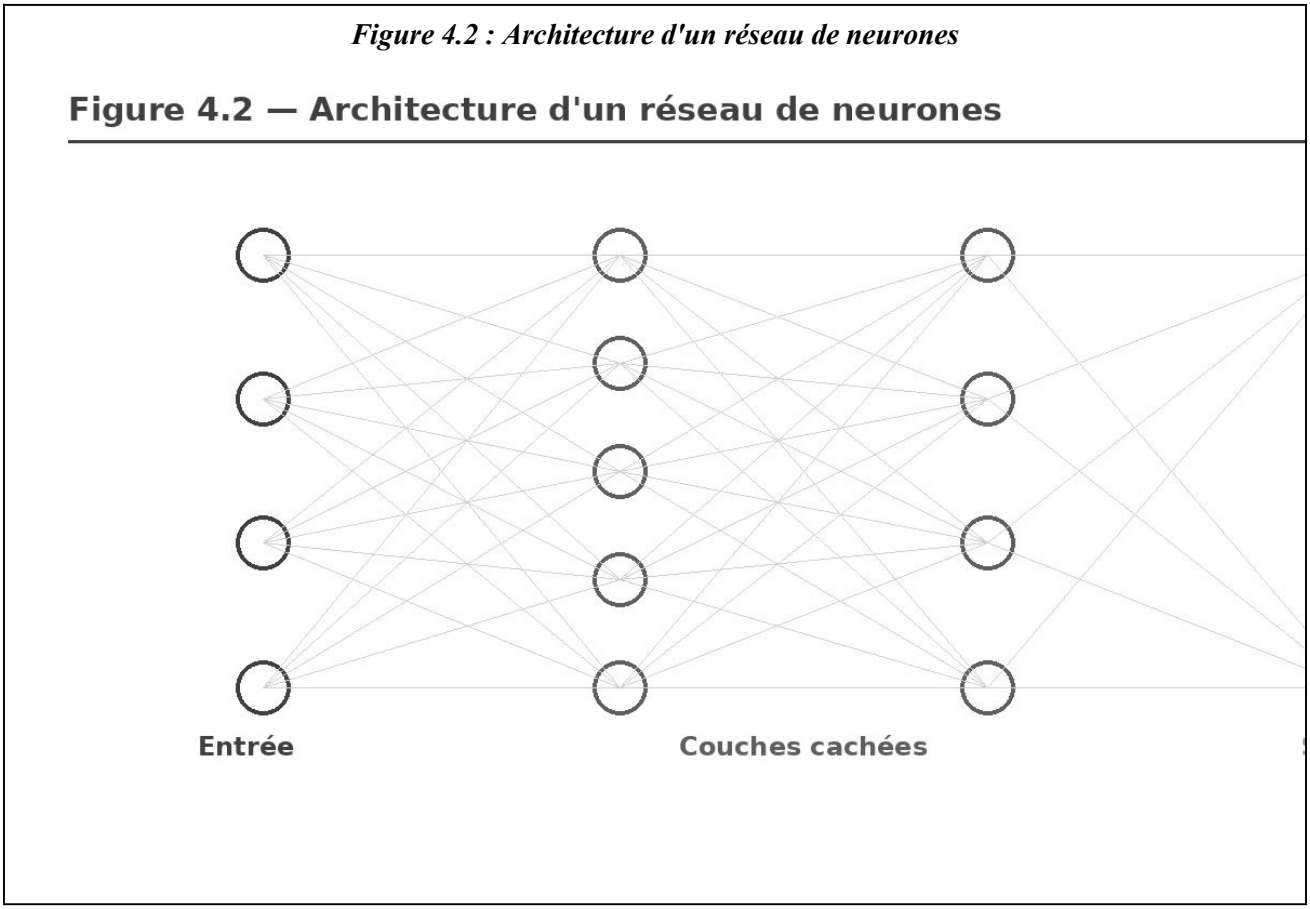
Analogie : Le jury de dégustation

Imaginez un jury de cinq experts qui goûtent un plat. Chaque expert donne une note, mais leur avis n'a pas le même poids : l'expert en gastronomie compte double, l'amateur une fois. La note finale résulte de cette somme pondérée. Si la note dépasse un seuil, le plat est sélectionné. C'est exactement ce que fait un neurone artificiel : somme pondérée des entrées, comparaison à un seuil, décision.

4.2 Le réseau de neurones des couches qui se spécialisent

Un réseau de neurones est constitué de plusieurs couches de neurones artificiels. La couche d'entrée reçoit les données brutes. Les couches cachées (hidden layers) transforment progressivement les données. La couche de sortie produit le résultat final. C'est le nombre de couches cachées qui détermine la « profondeur » du réseau : d'où l'appellation Deep Learning pour les réseaux avec de nombreuses couches.

Figure 4.2 : Architecture d'un réseau de neurones



4.2.1 La rétropropagation : comment le réseau apprend

L'algorithme qui permet au réseau d'apprendre s'appelle la rétropropagation (backpropagation). Son fonctionnement, simplifié, se déroule en deux passes. La passe avant : les données entrent, traversent le réseau de gauche à droite, et produisent une prédiction. La passe arrière : on calcule l'erreur entre la prédiction et la réalité, puis on remonte cette erreur de droite à gauche dans le réseau, en ajustant les poids de chaque neurone pour réduire cette erreur. Ce cycle se répète jusqu'à convergence [1].

4.3 Les architectures spécialisées

Au fil des années, des architectures spécialisées ont été développées pour différents types de données.

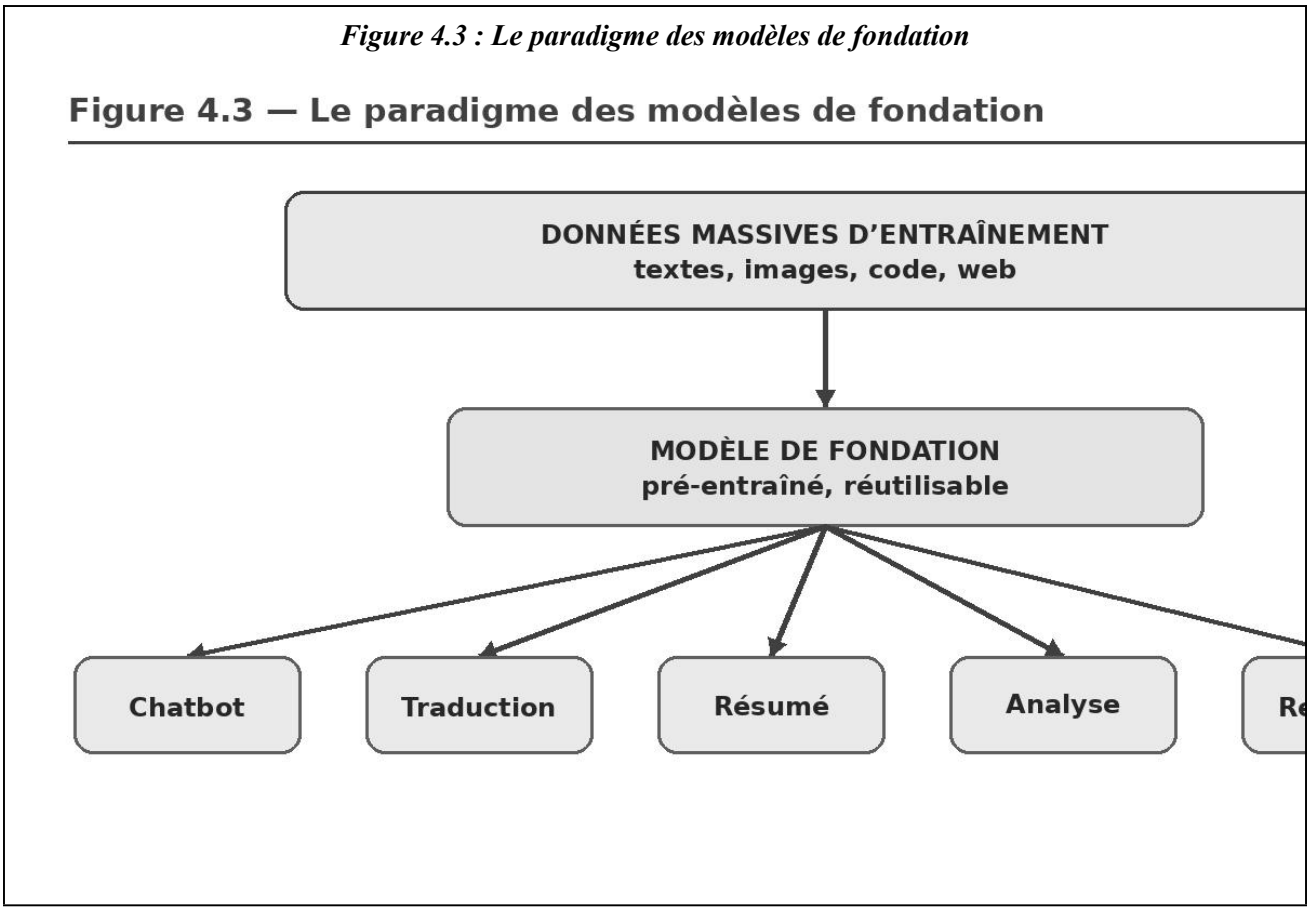
Architecture	Spécialité	Principe clé	Applications
CNN (Réseau convolutif)	Images	Détecte des patterns locaux (bords, formes)	Reconnaissance d'images, vidéos, imagerie médicale
RNN / LSTM	Séquences temporelles	Mémorise le contexte précédent	Traduction, prévision, transcription
Transformer	Texte et multimodal	Attention : chaque mot regarde tous les autres	ChatGPT, Claude, Gemini, traduction, résumé
GAN	Génération	Deux réseaux s'affrontent : générateur vs discriminateur	Images synthétiques, deepfakes, augmentation de données
Diffusion	Génération d'images	Dé-bruitage progressif d'un bruit aléatoire	DALL-E, Midjourney, Stable Diffusion

Tableau 4.1 : Architectures principales de réseaux de neurones. Source : adapté de [2].

4.4 Les modèles de fondation : une révolution d'échelle

Depuis 2020, une nouvelle catégorie de modèles a émergé : les modèles de fondation (foundation models) [3]. Ces modèles sont pré-entraînés sur des quantités astronomiques de données des centaines de milliards de mots, des milliards d'images : puis adaptés (fine-tuned) pour des tâches spécifiques. Ce paradigme change tout au lieu de construire un modèle spécifique pour chaque tâche, on adapte un modèle général déjà très capable.

Figure 4.3 : Le paradigme des modèles de fondation



4.4.1 Les principaux modèles de 2025-2026

Modèle	Créateur	Points forts	Accès
GPT-4o	OpenAI (USA)	Multimodal, très performant en langage et image	Payant (API + ChatGPT)
Claude 3.5 / 4	Anthropic (USA)	Contexte long, rigueur factuelle, sécurité	Payant / API
Gemini 1.5 / 2	Google DeepMind	Intégration Google Workspace, multimodal	Gratuit (limité) + API
Llama 3.x	Meta (USA)	Open source, déployable localement	Gratuit (open source)
DeepSeek R1 / V3	DeepSeek (Chine)	Très performant, coût d'entraînement réduit	Gratuit / API
Falcon 2	TII (Émirats)	Arabe et multilingue, open source	Gratuit (open source)
Jais	MBZUAI (Émirats)	Spécialisé arabe, contexte régional	Disponible via API

Tableau 4.2 : Principaux modèles de fondation disponibles en 2025-2026.

À retenir sur le Chapitre 4

Un neurone artificiel fait une somme pondérée de ses entrées et décide si le signal passe selon une fonction d'activation. Un réseau de neurones profond apprend des représentations hiérarchiques : chaque couche extrait des caractéristiques de plus en plus abstraites. La rétropropagation ajuste les poids du réseau en remontant l'erreur de la sortie vers l'entrée. Les architectures se spécialisent selon le type de données : CNN pour les images, Transformer pour le texte. Les modèles de fondation sont pré-entraînés sur des données massives et ensuite adaptés à des tâches spécifiques.

Références bibliographiques : Chapitre 4

Réseaux de neurones et modèles de fondation

Références citées

[1] Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). « Learning representations by back-propagating errors ». Nature, vol. 323, n° 6088, p. 533-536. DOI : 10.1038/323533a0.

[2] Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. Cambridge : MIT Press. ISBN 978-0-262-03561-3. https://www.deeplearningbook.org.

[3] Bommasani, R. et al. (2021). « On the Opportunities and Risks of Foundation Models ». arXiv : 2108.07258. DOI : 10.48550/arXiv.2108.07258.

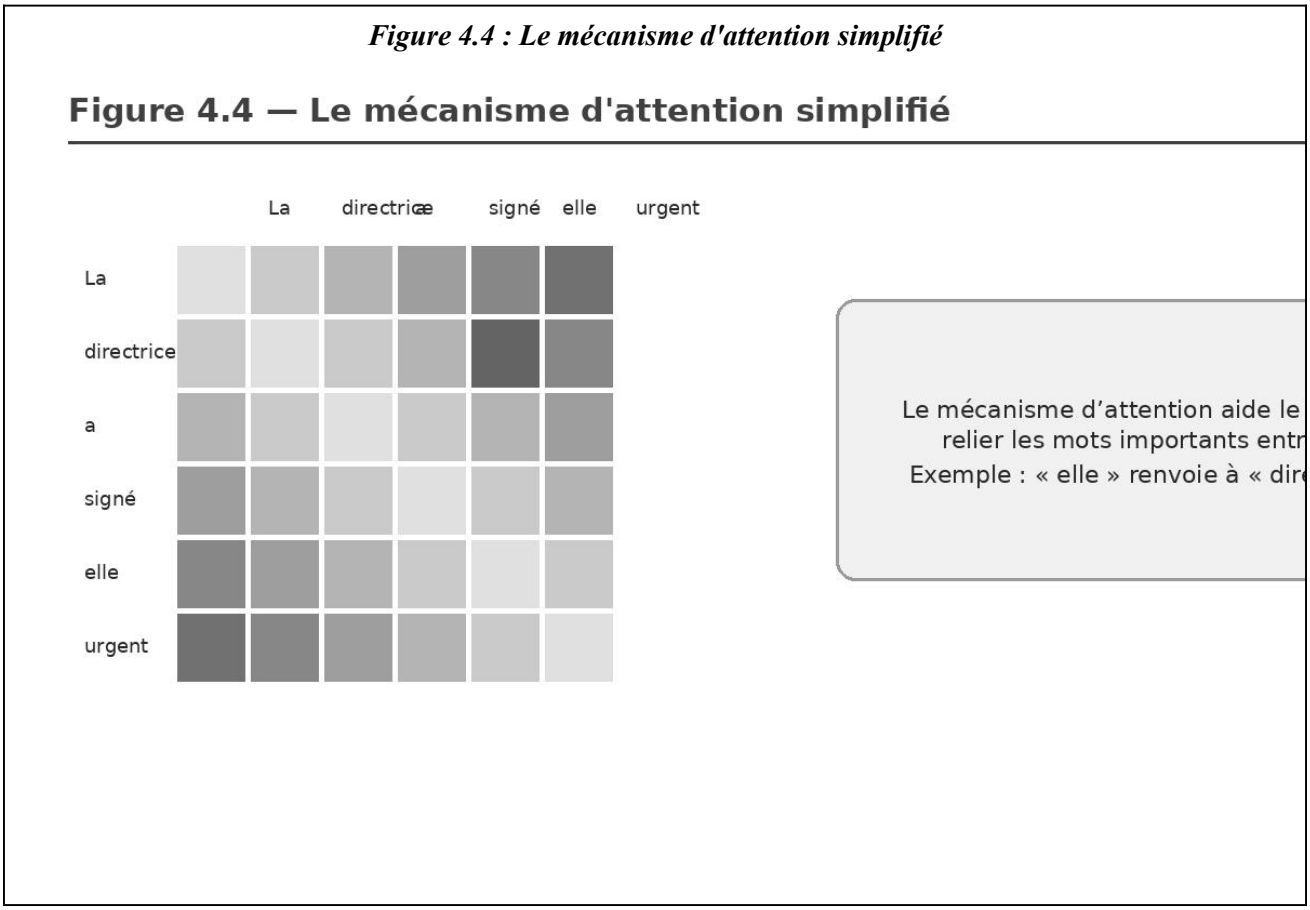
Extension Chapitre 4 : Les Transformers en détail

4.5 Le mécanisme d'attention : cœur des Transformers

L'innovation centrale du Transformer est le mécanisme d'attention multi-têtes (multi-head attention). L'idée fondamentale est que, pour traiter un mot dans une phrase, le modèle doit évaluer l'importance de chaque autre mot dans la phrase. Ce mécanisme se déroule en trois étapes que l'on peut expliquer intuitivement.

Chaque token est transformé en trois représentations différentes : une clé (Key), une valeur (Value) et une requête (Query). Pour calculer comment le mot « elle » dans la phrase « La directrice a signé son rapport car elle l'estimait urgent » est lié aux autres mots, le mécanisme compare la requête de « elle » avec les clés de tous les autres mots. Si la comparaison entre « elle » et « directrice » donne un score élevé, le modèle attribue un poids important à la valeur de « directrice » pour interpréter « elle ». Ce calcul se fait en parallèle pour tous les mots de la phrase : d'où l'efficacité du Transformer sur des textes longs [1].

Figure 4.4 : Le mécanisme d'attention simplifié



4.6 Fine-tuning vs zero-shot vs few-shot

Quand on utilise un modèle de fondation pour une nouvelle tâche, trois approches se distinguent, avec des compromis différents en coût, performance et données requises.

Approche	Principe	Données requises	Coût	Performance
Fine-tuning	Réentraîner le modèle sur des données spécifiques à la tâche	Des centaines à milliers d'exemples étiquetés	Élevé (GPU, temps)	Très bonne sur la tâche spécifique
Few-shot	Fournir quelques exemples dans le prompt sans réentraîner	2 à 10 exemples dans le prompt	Nul	Bonne si exemples bien choisis
Zero-shot	Décrire la tâche en langage naturel sans exemple	Aucune donnée supplémentaire	Nul	Variable selon la tâche

Tableau 4.3 : Comparaison des approches d'utilisation d'un modèle de fondation.

Exemple : Trois façons d'utiliser un LLM pour analyser un discours politique

- Fine-tuning : entraîner un modèle sur 5000 discours politiques algériens étiquetés selon leur registre (populiste, technocratique, patriotique...). Le modèle apprend précisément les codes rhétoriques locaux. Coût : plusieurs semaines de travail, GPU, expertise.
- Few-shot : dans le prompt, donner 3 exemples de discours avec leur étiquette, puis demander au modèle de classer un nouveau discours. Coût : quelques secondes. Performance moindre mais souvent suffisante.
- Zero-shot : demander directement 'Classifie ce discours politique selon son registre rhétorique dominant'. Fonctionne bien pour les grandes langues, moins bien pour l'arabe dialectal algérien.

4.7 L'évaluation des modèles de langage : les benchmarks

Les performances des LLM sont comparées sur des benchmarks standardisés. Les connaître permet d'interpréter les annonces de performances spectaculaires avec le recul nécessaire.

Benchmark	Ce qu'il mesure	Limite connue
MMLU	Connaissances académiques dans 57 domaines (questions à choix multiple)	Mémorise les réponses si présentes dans les données d'entraînement
HumanEval	Capacité à générer du code correct	Ne mesure que des tâches de code simples et bien définies
TruthfulQA	Tendance à produire des réponses vraies vs populaires	Basé sur des questions anglophones, peu adapté aux contextes arabes
ArabicMMLU	Connaissances en arabe dans plusieurs domaines	En cours de développement, couverture dialectale limitée
BIG-Bench Hard	Raisonnement complexe et tâches	Certaines tâches peuvent être résolues par

Benchmark	Ce qu'il mesure	Limite connue
	difficiles	mémorisation

Tableau 4.4 : Principaux benchmarks d'évaluation des LLM et leurs limites.

Références bibliographiques : Chapitre 4-Extension

Les Transformers en détail

Références

[1] Vaswani, A. et al. (2017). « Attention Is All You Need ». Advances in Neural Information Processing Systems, vol. 30. DOI : 10.48550/arXiv.1706.03762.

Chapitre 5 : Traitement du langage naturel et IA générative

Objectifs pédagogiques

- Comprendre comment une machine traite et génère du langage humain.
- Saisir le fonctionnement des modèles de langage (LLM) mot par mot.
- Comprendre ce qu'est l'IA générative et comment elle produit du contenu.
- Identifier les usages et limites du traitement automatique du langage arabe.
- Comprendre pourquoi les modèles de langage produisent des hallucinations.

5.1 Le traitement automatique du langage (TAL / NLP)

Le traitement automatique du langage naturel (TAL en français, NLP en anglais) est la branche de l'IA qui s'occupe du langage humain : textes, paroles, dialogues. C'est l'une des tâches les plus difficiles en IA, car le langage est ambigu, contextuel, culturel et infiniment variable.

5.1.1 De la phrase au vecteur : comment une machine représente les mots

Une machine ne peut pas traiter directement des mots : elle ne comprend que des nombres. Pour traiter du texte, les systèmes d'IA convertissent d'abord chaque mot (ou sous-mot) en un vecteur numérique, c'est-à-dire une liste de nombres. Ces vecteurs sont construits de manière à ce que des mots de sens proche aient des vecteurs proches dans l'espace mathématique. C'est ce qu'on appelle les embeddings [1].

Exemple : Les embeddings

Après entraînement, dans un espace d'embeddings bien construit : le vecteur de 'roi' moins le vecteur de 'homme' plus le vecteur de 'femme' donne un vecteur très proche de 'reine'. Le système n'a pas appris cette règle : il l'a déduite des patterns du langage en lisant des milliards de phrases.

5.1.2 L'attention : le mécanisme clé des Transformers

Le Transformer, l'architecture qui propulse tous les grands modèles de langage, repose sur un mécanisme appelé l'attention (attention mechanism) [2]. L'idée est simple : pour comprendre un mot dans une phrase, il faut regarder tous les autres mots de la phrase et évaluer lequel est le plus utile pour interpréter ce mot.

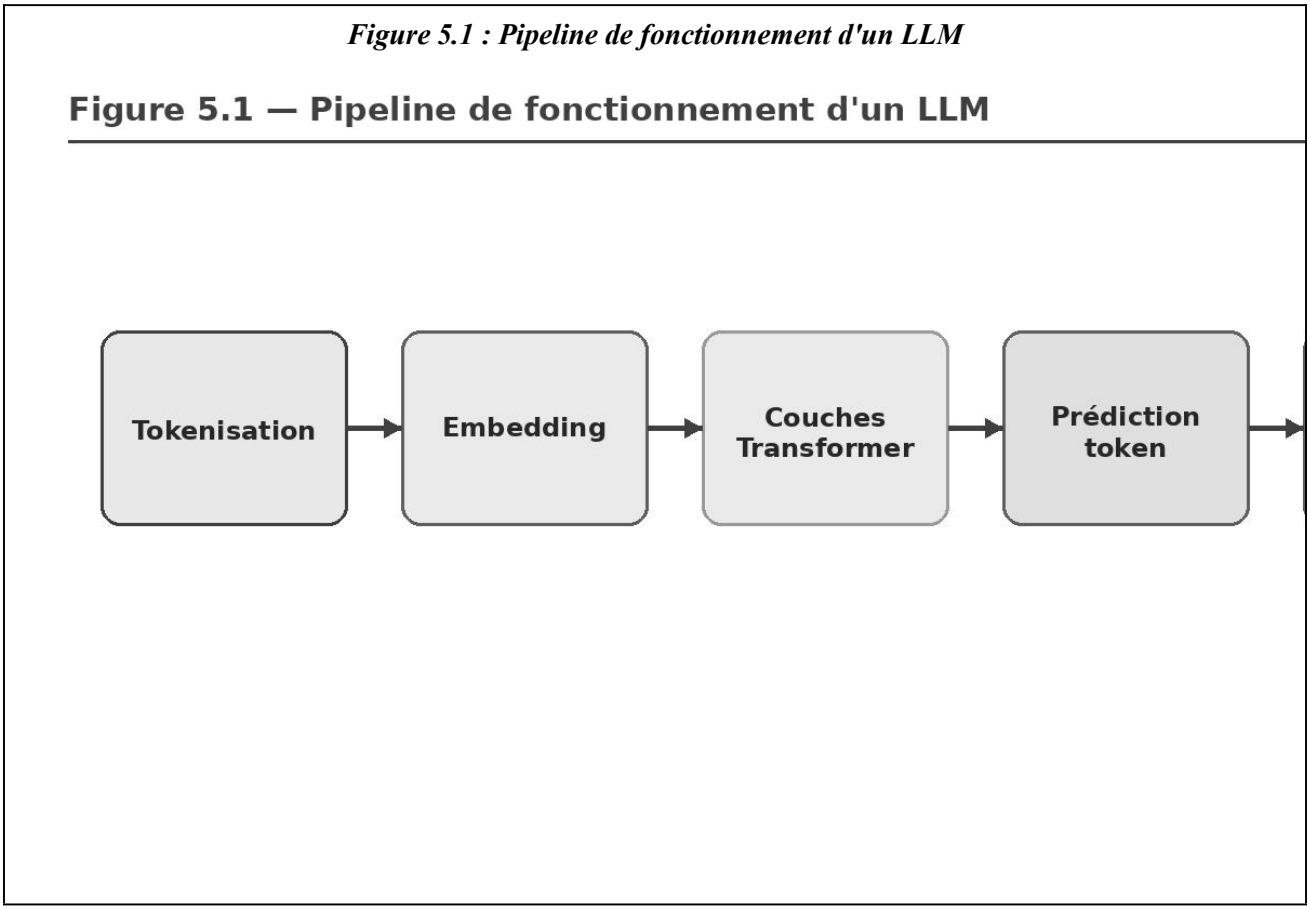
Exemple : Ambiguïté et contexte

Dans la phrase 'La banque a refusé mon prêt car elle était en faillite', à quoi réfère 'elle' ? À 'la banque', bien sûr, pas à 'ma demande'. L'attention permet au modèle d'évaluer que 'elle' est fortement lié à 'banque' dans ce contexte. C'est ce mécanisme qui rend les Transformers si efficaces pour le langage.

5.2 Comment fonctionne un modèle de langage (LLM)

Un grand modèle de langage (Large Language Model, LLM) est entraîné à prédire le mot suivant dans une phrase, encore et encore, sur des centaines de milliards de phrases. C'est un objectif d'apprentissage en apparence simple, mais qui nécessite : pour être bien réalisé : une compréhension implicite de la grammaire, des faits, de la logique, du style et des conventions culturelles.

Figure 5.1 : Pipeline de fonctionnement d'un LLM



5.2.1 Pourquoi ChatGPT invente parfois des faits

Un modèle de langage ne cherche pas la vérité : il cherche la cohérence. Il prédit le mot qui est le plus probable dans le contexte donné, pas nécessairement le mot le plus exact. Si la question porte sur un sujet peu représenté dans ses données d'entraînement, il peut générer une réponse qui semble correcte stylistiquement mais qui est factuellement fausse. Ce phénomène s'appelle l'hallucination [3] : nous y reviendrons en détail dans le chapitre 7.

5.3 L'IA générative

L'IA générative désigne les systèmes capables de produire du contenu nouveau : textes, images, sons, vidéos, code. Contrairement à l'IA discriminative qui classe des données existantes (spam ou pas spam), l'IA générative crée. Cette distinction est fondamentale pour comprendre les nouveaux usages et les nouveaux risques.

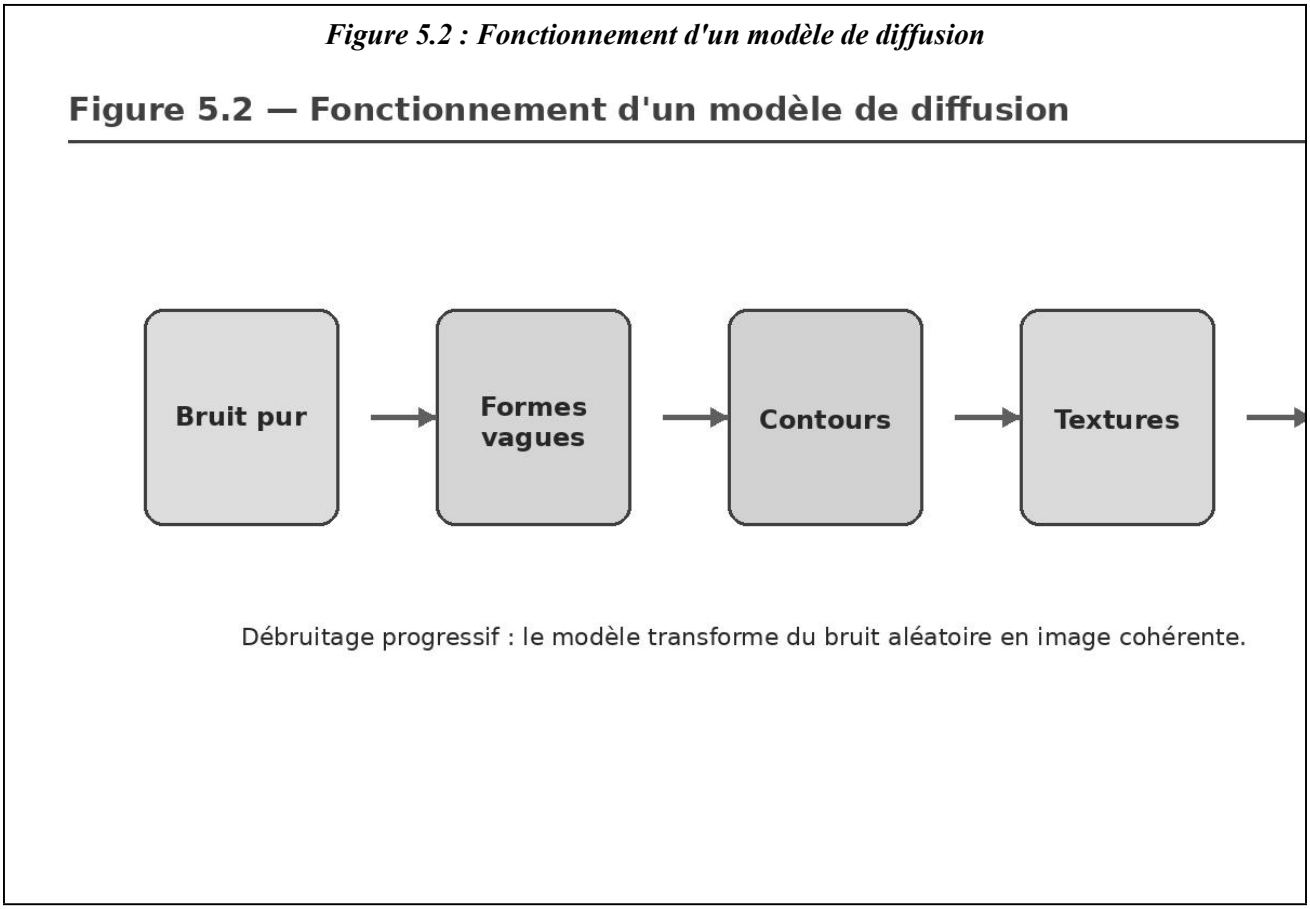
Dimension	IA discriminative	IA générative
Objectif	Classer, détecter, prédire	Créer, générer, produire
Sortie	Étiquette, score, catégorie	Texte, image, son, vidéo, code
Exemples	Filtre anti-spam, détection de fraude, diagnostic médical	ChatGPT, DALL-E, Midjourney, Sora, GitHub Copilot
Risques spécifiques	Biais, discrimination, opacité	Désinformation, deepfakes, plagiat, contrefaçon

Tableau 5.1 : IA discriminative vs IA générative.

5.3.1 Comment DALL-E génère une image

Les modèles génératifs d'images les plus récents (DALL-E 3, Midjourney, Stable Diffusion) utilisent une architecture appelée modèle de diffusion. Le principe : on part d'une image de bruit aléatoire (comme de la neige sur un écran) et on la dé-bruite progressivement, guidé par le texte de description. À chaque étape, le modèle retire un peu de bruit et ajoute un peu de cohérence, jusqu'à obtenir une image qui correspond au texte.

Figure 5.2 : Fonctionnement d'un modèle de diffusion



5.3.2 Le traitement du langage arabe : défis et avancées

La langue arabe pose des défis spécifiques au traitement automatique : la morphologie riche (un mot peut avoir des dizaines de formes), la polysémie, les dialectes très variés (darija algérienne, égyptien, golfe...), et le manque de données annotées de qualité. Les modèles entraînés sur des corpus majoritairement anglais performent nettement moins bien sur l'arabe et encore moins sur les dialectes [4].

Des initiatives prometteuses existent. AraBERT [5], développé par l'Université arabe de Beyrouth, et Jais, développé par le Mohamed bin Zayed University of AI (MBZUAI) aux Émirats arabes unis, proposent des modèles spécifiquement entraînés sur des données arabes. Ces ressources sont cruciales pour l'Algérie, où le traitement automatique de la darija reste largement sous-développé.

À retenir sur le Chapitre 5

Le NLP transforme le langage en nombres (embeddings) pour que la machine puisse le traiter.

Le mécanisme d'attention permet à chaque mot d'évaluer sa relation avec tous les autres dans la phrase.

Un LLM génère du texte mot après mot, en prédisant à chaque étape le token le plus probable.

L'IA générative crée du contenu nouveau (texte, image, son) à la différence de l'IA discriminative qui classe.

L'arabe, et plus encore la darija, sont sous-représentés dans les corpus d'entraînement des grands modèles.

Activité 5.1 : Expérimentation critique

Posez à un LLM disponible (ChatGPT, Gemini ou autre) les questions suivantes, et analysez les réponses :

a) Demandez-lui de résumer un événement politique algérien récent. Vérifiez les faits annoncés.

b) Demandez-lui de traduire une phrase en darija algérienne. Évaluez la qualité.

c) Demandez-lui d'expliquer comment il fonctionne. Analysez si l'explication est exacte au regard de ce chapitre.

Pour chaque réponse, identifiez les forces, les limites et les hallucinations éventuelles.

Références bibliographiques : Chapitre 5

Traitement du langage naturel et IA générative

Références citées

[1] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). « Distributed Representations of Words and Phrases and their Compositionality ». Advances in Neural Information Processing Systems, vol. 26. arXiv : 1310.4546. DOI : 10.48550/arXiv.1310.4546.

[2] Vaswani, A. et al. (2017). « Attention Is All You Need ». Advances in Neural Information Processing Systems, vol. 30. arXiv : 1706.03762. DOI : 10.48550/arXiv.1706.03762.

[3] Ji, Z. et al. (2023). « Survey of Hallucination in Natural Language Generation ». ACM Computing Surveys, vol. 55, n° 12, art. 248. DOI : 10.1145/3571730.

[4] Darwish, K. & Magdy, W. (2014). « Arabic Information Retrieval ». Foundations and Trends in Information Retrieval, vol. 7, n° 4, p. 239-342. DOI : 10.1561/1500000031.

[5] Antoun, W., Baly, F. & Hajj, H. (2020). « AraBERT : Transformer-based Model for Arabic Language Understanding ». Proceedings of the 4th Workshop on Open-Source Arabic Corpora, LREC 2020. arXiv : 2003.00104. DOI : 10.48550/arXiv.2003.00104.

Complément du Chapitre 5 : Comparatif des modèles génératifs et limites

5.4 Comparatif : IA classique versus IA générative

La distinction entre IA discriminative et IA générative est fondamentale pour comprendre les nouveaux risques. Ce tableau comparatif en offre une synthèse accessible.

Critère	IA discriminative	IA générative
Objectif principal	Classer, détecter, prédire une catégorie	Créer un contenu nouveau
Mode de sortie	Étiquette, score, probabilité	Texte, image, audio, vidéo, code
Données nécessaires	Étiquetées (supervisé), souvent moins massives	Très massives, souvent non étiquetées
Exemples (2025)	Détecteur de spam, classifieur médical, OCR	ChatGPT, DALL-E 3, Midjourney, Sora, Copilot
Atouts	Très précis sur la tâche définie, interprétable	Créatif, polyvalent, économique d'emploi
Risques spécifiques	Biais, discrimination, opacité des décisions	Désinformation, deepfakes, hallucinations, dépendance
Régulation actuelle	Couverte par le RGPD, AI Act (applications à risque)	Nouveaux cadres en construction (marquage, watermark)

Tableau 5.2 : Comparatif IA discriminative vs IA générative.

5.5 Les limites techniques actuelles des LLM

Les grands modèles de langage sont impressionnants, mais leurs limites sont tout aussi importantes à connaître que leurs capacités. Six limites méritent une attention particulière de la part du politiste.

5.5.1 La coupure temporelle

Tout LLM a une date de coupure des données d'entraînement (training cutoff). Après cette date, il n'a aucune information sur le monde. Si vous demandez à un modèle de 2024 de commenter un événement de 2025, il inventera ou refusera. Certains modèles intègrent des outils de recherche Web pour pallier cette limite (mode avec accès Internet), mais cela reste une couche externe, pas une compréhension native des événements récents.

5.5.2 L'absence de mémoire entre sessions

Par défaut, un LLM ne se souvient pas de vos conversations précédentes. Chaque session commence à zéro, avec uniquement le contexte de la conversation en cours. Des solutions d'ajout de mémoire externe existent (bases de données vectorielles, RAG), mais ce n'est pas une mémoire organique comme celle d'un humain.

5.5.3 Le contexte limité

Un LLM peut traiter un nombre maximum de tokens simultanément : sa « fenêtre de contexte ». Les meilleurs modèles actuels atteignent un million de tokens, soit environ 750 000 mots. Au-delà, le modèle ne peut pas analyser le document entier en une seule passe. Les textes ou conversations très longs exigent des stratégies de découpage.

5.5.4 L'incapacité à raisonner vraiment

Les LLM peuvent sembler raisonner : enchaîner des étapes logiques, résoudre des problèmes complexes. Mais ce qu'ils font est fondamentalement différent du raisonnement humain : ils génèrent la suite la plus probable de tokens, pas nécessairement la suite la plus logique. Sur des problèmes de logique formelle inhabituels ou sur des raisonnements en plusieurs étapes, ils échouent souvent de manière spectaculaire et avec la même confiance que quand ils réussissent.

5.5.5 L'opacité et la non-explicabilité

On ne peut pas demander à un LLM d'expliquer précisément pourquoi il a produit telle réponse plutôt qu'une autre. Les techniques d'explicabilité (XAI) permettent d'identifier quels tokens d'entrée ont eu le plus d'influence, mais elles ne fournissent pas une explication causale complète. Cette opacité est problématique pour les usages dans des domaines où la justification des décisions est légalement requise.

5.5.6 Le coût environnemental

L'entraînement de grands modèles consomme des quantités massives d'énergie. L'inférence : c'est-à-dire répondre à une question : consomme beaucoup moins, mais à l'échelle de milliards de requêtes quotidiennes, le bilan reste significatif. L'IEA (Agence internationale de l'énergie) a publié en 2024 des projections alarmantes sur la croissance de la consommation électrique des datacenters d'IA [1].

Références bibliographiques : Chapitre 5-Complément

Comparatif et limites des modèles génératifs

Références complémentaires

[1] International Energy Agency (2024). Electricity 2024 : Analysis and Forecast to 2026. Paris : IEA. https://www.iea.org/reports/electricity-2024.

Chapitre 6 : Vision par ordinateur, systèmes de recommandation et chatbots

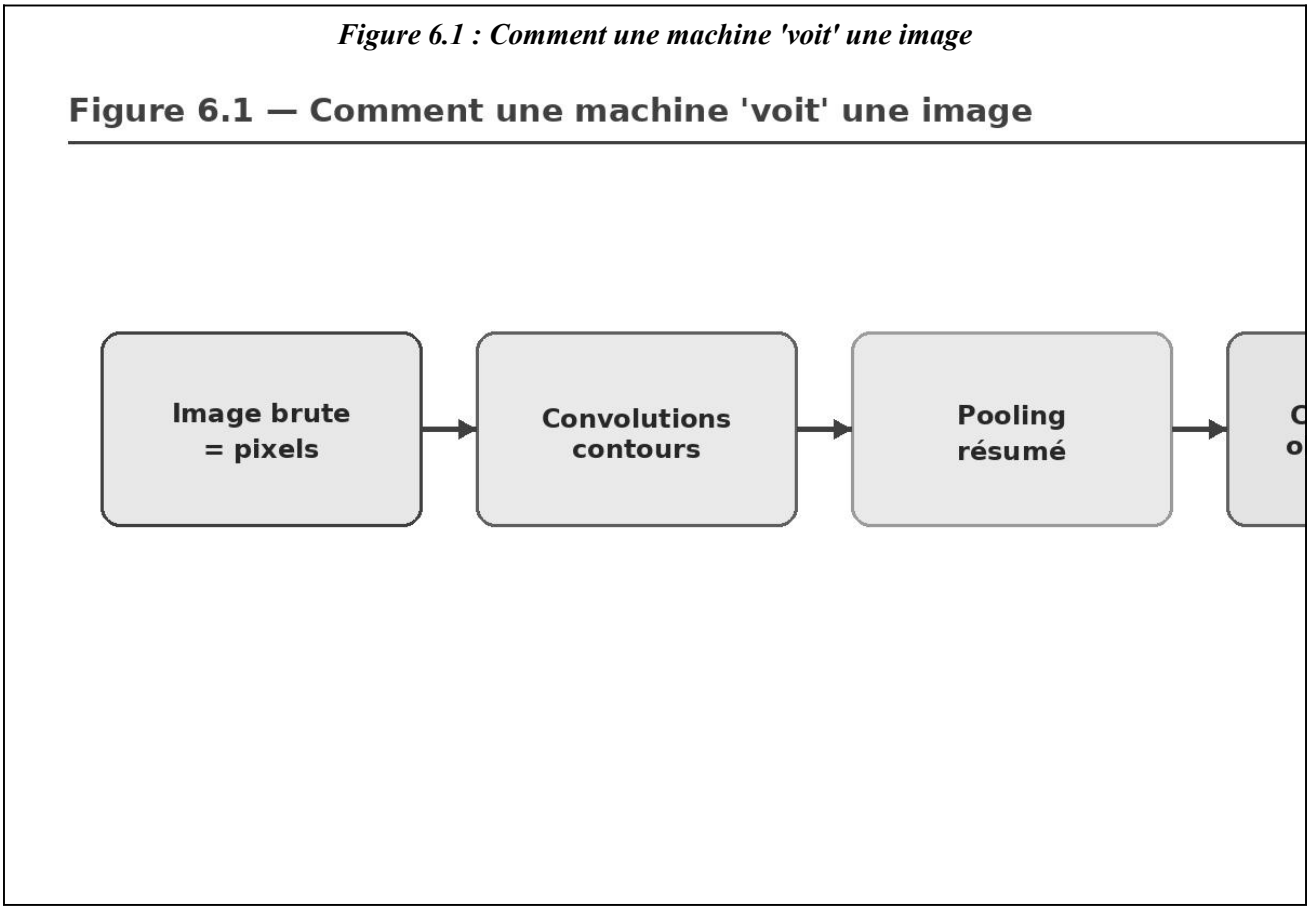
Objectifs pédagogiques

- Comprendre comment une machine 'voit' et interprète une image.
- Saisir le fonctionnement d'un système de recommandation.
- Comprendre comment un chatbot génère une réponse.
- Identifier les applications politiques, administratives et sociales de ces technologies.
- Distinguer ce qui est techniquement possible de ce qui est réellement déployé.

6.1 La vision par ordinateur

Pour un humain, reconnaître un visage ou lire un panneau routier est instantané et naturel. Pour une machine, c'est un défi mathématique complexe. Une image numérique n'est, pour un ordinateur, qu'un tableau de nombres : chaque pixel est représenté par trois valeurs (rouge, vert, bleu, de 0 à 255). La vision par ordinateur consiste à transformer ces tableaux de nombres en informations utiles, qui est sur cette photo, qu'est-ce que ce texte, quel objet est-ce.

Figure 6.1 : Comment une machine 'voit' une image



6.1.1 La reconnaissance faciale : une technologie politique

La reconnaissance faciale est l'une des applications de vision par ordinateur les plus répandues et les plus controversées. Elle identifie ou vérifie l'identité d'une personne à partir de son image. Les systèmes les plus performants atteignent des taux d'exactitude supérieurs à 99 % sur des visages bien éclairés d'adultes aux profils bien représentés dans les données d'entraînement.

Mais ces systèmes présentent des biais documentés : une étude publiée dans Proceedings of Machine Learning Research [1] a montré que les taux d'erreur varient considérablement selon le genre et la couleur de peau : les femmes à peau foncée étant les plus mal reconnues. Ces biais ont des conséquences réelles des cas d'arrestations erronées aux États-Unis ont été directement attribués à des identifications faciales incorrectes.

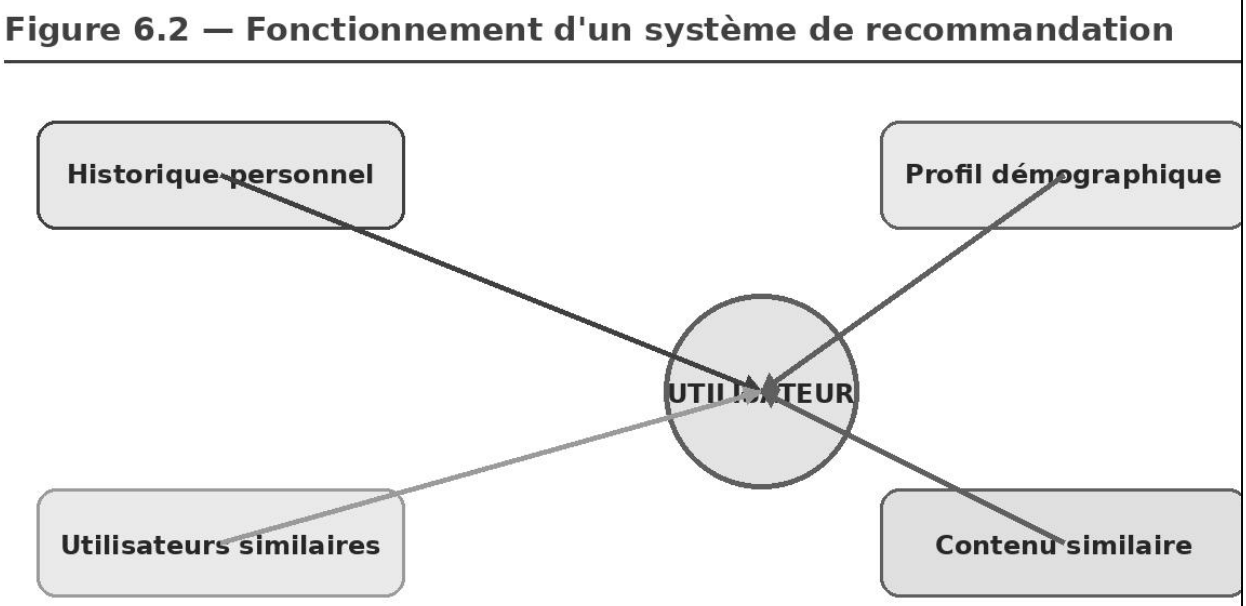
6.1.2 Vision par ordinateur dans la gouvernance

Les applications gouvernementales de la vision par ordinateur sont nombreuses : contrôle aux frontières, surveillance urbaine, lecture automatique de plaques d'immatriculation, analyse de documents administratifs (reconnaissance optique de caractères), suivi de trafic, détection d'anomalies dans des images satellites. En Algérie, la carte d'identité biométrique utilise déjà des techniques de vision par ordinateur pour l'identification sécurisée.

6.2 Les systèmes de recommandation

Chaque fois que YouTube vous propose une vidéo, qu'Amazon vous suggère un produit, ou que Spotify sélectionne votre prochaine chanson, un système de recommandation est à l'œuvre. Ces systèmes sont parmi les applications d'IA les plus massives du monde en termes d'utilisation quotidienne.

Figure 6.2 : Fonctionnement d'un système de recommandation



6.2.1 La bulle de filtre : un enjeu démocratique

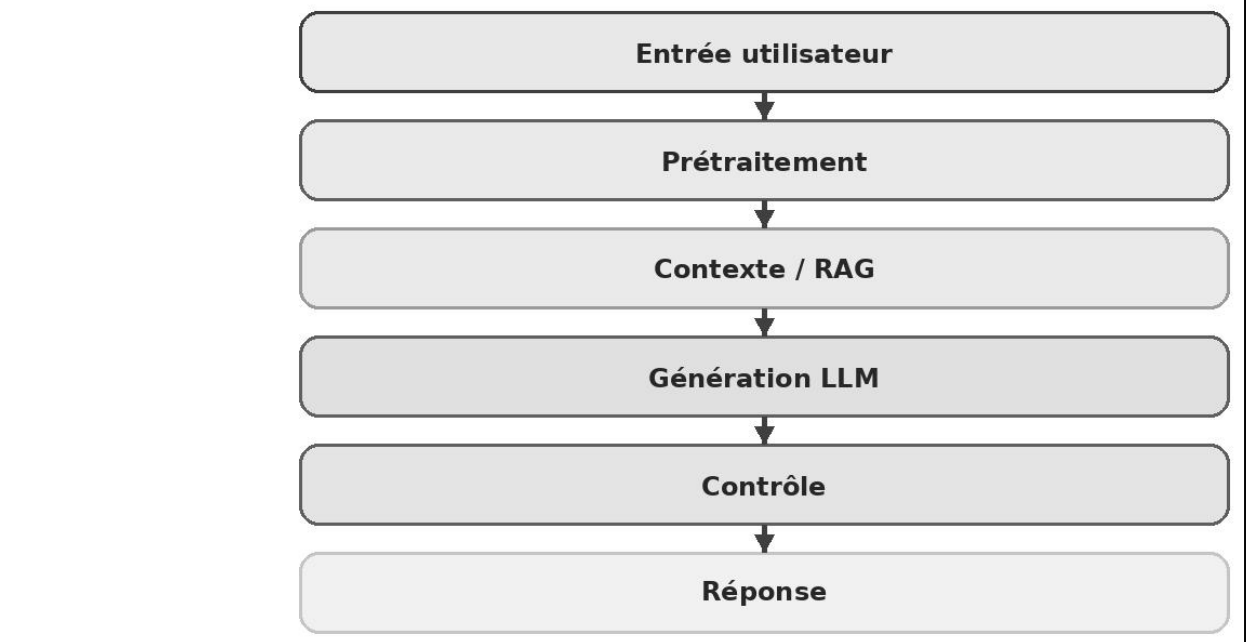
Les systèmes de recommandation optimisent l'engagement : le temps passé sur la plateforme. Ils apprennent que les contenus qui provoquent des réactions fortes (indignation, peur, surprise) génèrent plus de clics que les contenus nuancés. Ce biais d'optimisation peut conduire à une radicalisation progressive : l'algorithme montre des contenus de plus en plus extrêmes pour maintenir l'attention [2].

Eli Pariser a théorisé ce phénomène sous le nom de « filtre bulle » : chaque utilisateur reçoit une version personnalisée de la réalité, construite par l'algorithme, qui l'isole des points de vue différents [3]. Les conséquences politiques sont documentées : polarisation de l'opinion publique, diffusion des théories du complot, difficultés pour les partis modérés à atteindre les indécis.

6.3 Les chatbots : conversation avec une machine

Un chatbot est un programme qui simule une conversation humaine. Les chatbots actuels les plus avancés (ChatGPT, Claude, Gemini) reposent sur des LLM : nous avons vu comment ils fonctionnent au chapitre 5. Mais pour comprendre l'expérience utilisateur et les usages politiques, il faut ajouter quelques mécanismes supplémentaires.

Figure 6.3 : Pipeline complet d'un chatbot moderne



6.3.1 Usages politiques des chatbots

Les chatbots sont de plus en plus utilisés dans la sphère politique et administrative. Dans les administrations, ils répondent aux questions courantes des usagers (horaires, documents requis, procédures). Dans les campagnes électorales, ils simulent le dialogue avec le candidat pour toucher un maximum d'électeurs. Dans la désinformation, ils peuvent générer des millions de messages personnalisés ciblant des électeurs spécifiques : une forme de propagande scalable et peu coûteuse.

<u>À retenir sur le Chapitre 6</u>

La vision par ordinateur traite les images comme des tableaux de nombres et en extrait des informations (visages, objets, textes) via des réseaux convolutifs.

La reconnaissance faciale est performante mais biaisée : les taux d'erreur varient significativement selon la couleur de peau et le genre.

Les systèmes de recommandation optimisent l'engagement, pas la vérité, ce biais peut conduire à la polarisation.

Un chatbot moderne est un LLM augmenté de mécanismes de contexte, de modération et parfois de recherche externe (RAG).

Ces trois technologies ont des applications politiques directes qui justifient leur étude dans un cursus de sciences politiques.

Activité 6.1 : Analyse critique des recommandations

Pendant une semaine, observez votre fil d'actualité sur une plateforme de votre choix (YouTube, Instagram, TikTok, Facebook). Notez : (a) quel type de contenu vous est recommandé, (b) si des contenus politiques apparaissent spontanément, (c) si ces contenus semblent diversifiés ou orientés. À la fin de la semaine, rédigez une note d'une page sur l'effet que cet algorithme a pu avoir sur votre perception d'un sujet politique.

Références bibliographiques : Chapitre 6

Vision par ordinateur, systèmes de recommandation et chatbots

Références citées

[1] Buolamwini, J. & Gebru, T. (2018). « Gender Shades : Intersectional Accuracy Disparities in Commercial Gender Classification ». Proceedings of Machine Learning Research, vol. 81, p. 77-91. https://proceedings.mlr.press/v81/buolamwini18a.html.

[2] Huszár, F. et al. (2022). « Algorithmic amplification of politics on Twitter ». Proceedings of the National Academy of Sciences, vol. 119, n° 1, e2025334119. DOI : 10.1073/pnas.2025334119.

[3] Pariser, E. (2011). The Filter Bubble : What the Internet Is Hiding from You. New York : Penguin Press, 294 p. ISBN 978-1-59420-300-8.

Extension Chapitre 6 : Automatisation intelligente et RPA

6.4 La RPA augmentée par l'IA

La robotique logicielle (Robotic Process Automation, RPA) désigne l'automatisation de tâches répétitives réalisées sur ordinateur : saisie de données, copier-coller entre systèmes, envoi de courriels standards, extraction de documents. Contrairement à l'automatisation classique, la RPA augmentée par l'IA peut gérer des cas imprévus, lire des documents non structurés (PDF manuscrits, images) et s'adapter aux variations.

6.4.1 Applications dans l'administration algérienne

La RPA augmentée par l'IA offre des opportunités concrètes pour l'administration algérienne, particulièrement dans trois domaines où les volumes de traitement sont élevés et les tâches répétitives.

Le premier est le traitement des demandes d'actes administratifs. La saisie automatique des données depuis des formulaires scannés, la vérification de cohérence, l'envoi automatique de récépissés et de notifications sont des tâches qui peuvent être automatisées sans nécessiter de transformation profonde des processus existants.

Le deuxième est la gestion des déclarations fiscales. L'extraction automatique de données depuis les pièces justificatives soumises en ligne, le calcul des montants dus selon les règles en vigueur, et la détection des incohérences sont des candidats naturels à la RPA augmentée.

Le troisième est le tri et l'aiguillage des demandes citoyennes. Un système de traitement automatique du langage peut analyser les courriels entrants, les classer par thème et par urgence, et les aiguiller vers le service compétent sans remplacement du fonctionnaire, mais avec un gain de temps considérable.

6.5 Les systèmes de recommandation en détail

Nous avons présenté brièvement les systèmes de recommandation au chapitre 6. Approfondissons les deux approches principales, car elles ont des implications très différentes sur les biais qu'elles peuvent produire.

6.5.1 Le filtrage collaboratif

Le filtrage collaboratif (collaborative filtering) fonctionne sur le principe de la similarité entre utilisateurs : « Les personnes qui ont aimé les mêmes contenus que vous ont aussi aimé X ». Il ne nécessite pas de connaître le contenu en détail : il exploite la similarité des comportements. L'avantage est la découverte : vous pouvez être recommandé des contenus auxquels vous n'auriez pas pensé. L'inconvénient est le biais de popularité : les contenus populaires sont recommandés à tout le monde, les contenus de niche à personne. En politique, cela favorise les discours dominants au détriment des voix minoritaires.

6.5.2 Le filtrage basé sur le contenu

Le filtrage basé sur le contenu (content-based filtering) recommande des contenus similaires à ceux que vous avez aimés, en analysant les caractéristiques du contenu lui-même. Si vous avez regardé plusieurs documentaires sur la politique étrangère algérienne, le système recommandera d'autres documentaires sur ce thème. L'avantage est la précision sur vos goûts établis. L'inconvénient est le manque de sérendipité et l'enfermement dans une bulle thématique.

Les systèmes réels combinent les deux approches (filtrage hybride), mais les arbitrages entre exploitation des goûts connus et exploration de nouvelles frontières sont des choix de conception qui ont des effets politiques. Un système qui privilégie trop fortement l'exploitation crée des bulles de filtre; un système qui privilégie trop l'exploration dépayse l'utilisateur et réduit son engagement.

AXE 2

LIMITES, RISQUES ET ENJEUX ÉTHIQUES

المخاطر والقضايا الأخلاقية

Les fondements techniques étant posés, cet axe en explore les limites et les risques. Comprendre ce qu'une IA peut faire et ce qu'elle fait mal : est indispensable pour évaluer ses applications et concevoir des politiques publiques adaptées. Deux chapitres structurent cet axe : les erreurs, hallucinations et biais algorithmiques d'abord; puis la cybersécurité, l'automatisation intelligente et la souveraineté numérique.

Chapitre 7 : Erreurs, hallucinations et biais algorithmiques

Objectifs pédagogiques

- Distinguer erreur, biais et hallucination dans les systèmes d'IA.
- Comprendre les causes techniques des hallucinations dans les LLM.
- Identifier les trois sources principales de biais algorithmiques.
- Saisir les conséquences politiques et sociales des biais.
- Connaître les approches de détection et de mitigation des biais.

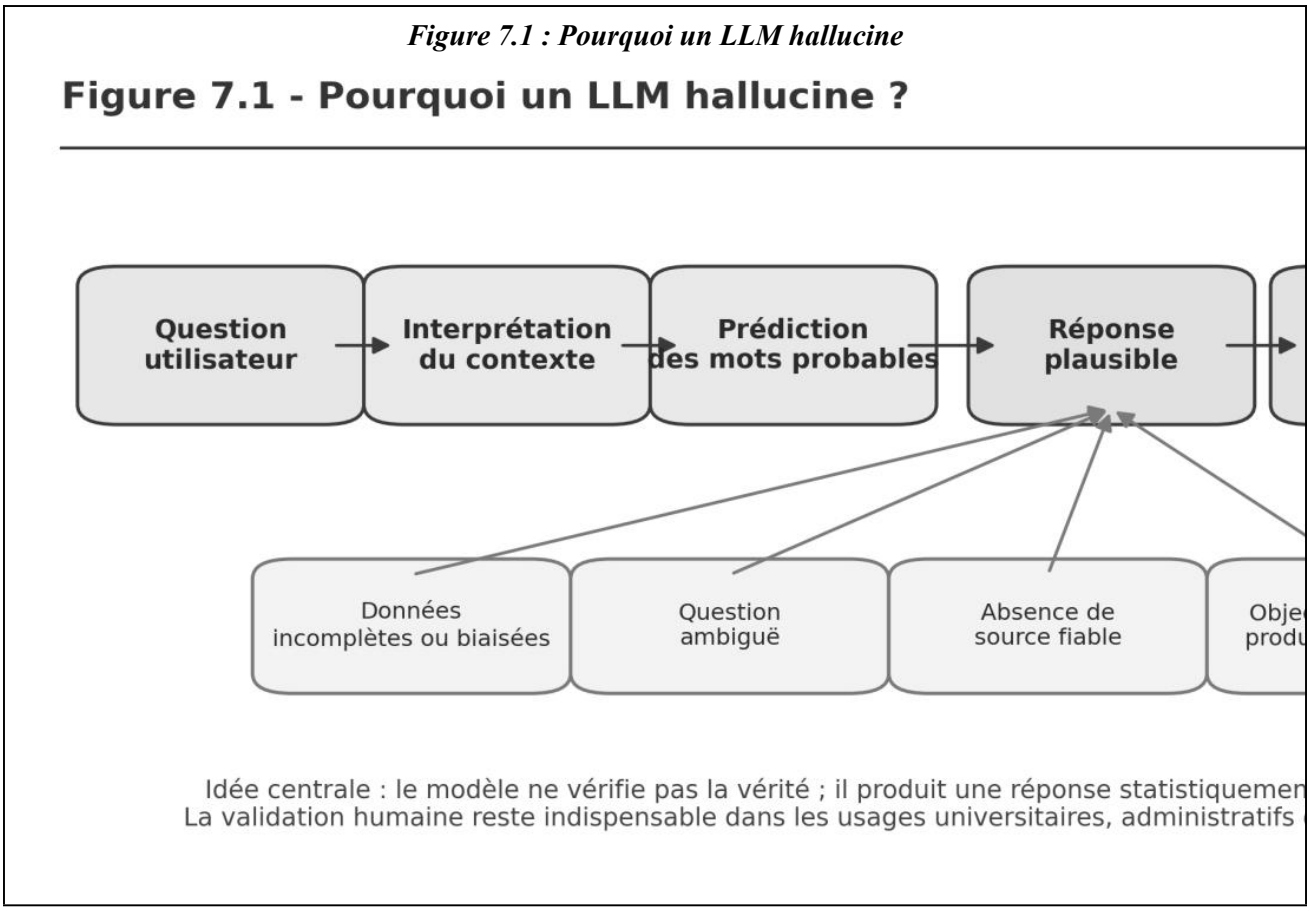
7.1 Les hallucinations des modèles de langage

En janvier 2023, un avocat américain a soumis à un tribunal un mémoire juridique contenant des citations d'arrêts de justice générées par ChatGPT. Ces arrêts n'existaient pas. Le modèle les avait inventés, avec leurs numéros de dossier, leurs dates, leurs juges et leurs attendus : plausibles en apparence, mais entièrement fictifs [1]. Le juge a sanctionné l'avocat pour imprudence. C'est un exemple d'hallucination : la production par un modèle d'une information fausse présentée avec assurance.

7.1.1 Pourquoi les modèles hallucinent

La cause technique est inhérente au fonctionnement même des LLM. Un modèle de langage génère le token le plus probable étant donné le contexte. Il n'a pas accès à une base de connaissances vérifiée : il a internalisé des patterns statistiques du langage. Lorsque la question porte sur un sujet rare dans ses données d'entraînement, il génère une réponse qui ressemble à une vraie réponse, car elle en a la forme et le style, mais dont le contenu est inventé. Le modèle n'a aucun mécanisme pour signaler son ignorance : il prédit toujours quelque chose.

Figure 7.1 : Pourquoi un LLM hallucine



7.1.2 Conséquences pratiques

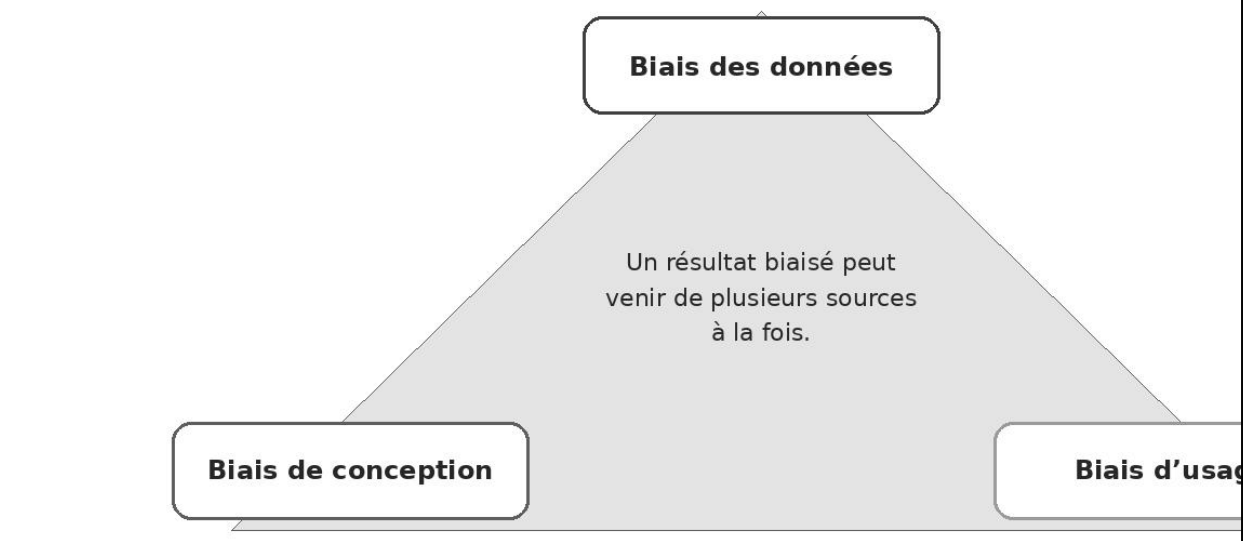
Les hallucinations posent des problèmes dans tous les domaines où la précision factuelle est critique : médecine, droit, histoire, diplomatie. Pour l'étudiant ou le chercheur en sciences politiques, la règle d'or est simple : ne jamais citer une information produite par un LLM sans l'avoir vérifiée dans une source primaire. Le LLM est un assistant de formulation et de structuration, pas une encyclopédie fiable.

7.2 Les biais algorithmiques

Un biais algorithmique est une distorsion systématique dans les prédictions ou décisions d'un système d'IA, qui avantage ou désavantage de manière injustifiée certains groupes. Les biais ne sont pas des bugs accidentels : ils reflètent des déséquilibres présents dans les données d'entraînement ou des choix de conception qui n'ont pas été suffisamment questionnés [2].

7.2.1 Trois sources de biais

Figure 7.2 : Les trois sources de biais algorithmique



7.2.2 L'affaire COMPAS : cas d'étude

ÉTUDE DE CAS : L'affaire COMPAS (2016)

COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) est un logiciel américain utilisé par des tribunaux pour estimer le risque de récidive d'un accusé et orienter les décisions de libération conditionnelle.

En 2016, l'organisation ProPublica a publié une analyse montrant que COMPAS classifiait incorrectement les accusés noirs comme à haut risque de récidive dans 45% des cas, contre 23% pour les accusés blancs. Inversement, les accusés blancs qui récidivèrent ensuite avaient été classés à faible risque dans 48% des cas, contre 28% pour les accusés noirs.

La cause : le modèle avait été entraîné sur des données historiques qui reflétaient les inégalités du système judiciaire américain : lui-même marqué par des décennies de discrimination raciale. L'algorithme avait appris et reproduit ces inégalités.

Références : Angwin, J. et al. (2016). « Machine Bias ». ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

7.2.3 Détecter et atténuer les biais

Approche	Description	Limite
Audit des données	Analyser la distribution des groupes dans les données d'entraînement	Ne détecte pas les biais implicites dans les étiquettes
Métriques d'équité	Mesurer les taux d'erreur par groupe (égalité d'opportunité, parité	Les définitions d'équité sont parfois mathématiquement incompatibles

Approche	Description	Limite
	démographique)	
Rééchantillonnage	Sur-représenter les groupes minoritaires dans l'entraînement	Peut réduire la performance globale
Explicabilité (XAI)	Ouvrir la boîte noire pour comprendre les décisions	Difficile à mettre en œuvre sur les grands réseaux
Révision humaine	Ajouter une validation humaine sur les cas sensibles	Scalabilité limitée, la revue humaine peut être biaisée aussi

Tableau 7.1 : Approches de détection et mitigation des biais. Source : adapté de [3].

À retenir : Chapitre 7

L'hallucination survient quand un LLM génère une information plausible en forme mais fausse en contenu. Cause : statistique, pas compréhension.

Un biais algorithmique est une distorsion systématique qui désavantage certains groupes. Ce n'est pas un bug : c'est souvent un reflet des données.

Les biais ont trois sources : les données (sous-représentation, étiquetage biaisé), la conception (choix de variables et de métriques), le déploiement (contexte inadapté).

L'affaire COMPAS illustre comment un algorithme peut amplifier des discriminations historiques systémiques.

Détecter et corriger les biais est difficile et les définitions mathématiques de l'équité sont parfois incompatibles entre elles.

Références bibliographiques : Chapitre 7

Erreurs, hallucinations et biais algorithmiques

Références citées

[1] Weiser, B. (2023). « Here's What Happens When Your Lawyer Uses ChatGPT ». The New York Times, 27 mai 2023. https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html.

[2] Mehrabi, N. et al. (2021). « A Survey on Bias and Fairness in Machine Learning ». ACM Computing Surveys, vol. 54, n° 6, art. 115. DOI : 10.1145/3457607.

[3] Bellamy, R. K. E. et al. (2019). « AI Fairness 360 : An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias ». IBM Journal of Research and Development, vol. 63, n° 4/5, art. 4:1. DOI : 10.1147/JRD.2019.2942287.

Complément du Chapitre 7 : Cadres éthiques et régulation internationale

7.3 Les principes éthiques internationaux

Face aux risques documentés de biais et de discrimination, plusieurs organisations internationales ont élaboré des principes éthiques pour guider le développement et le déploiement de l'IA. Connaître ces cadres est indispensable pour comprendre les débats réglementaires actuels.

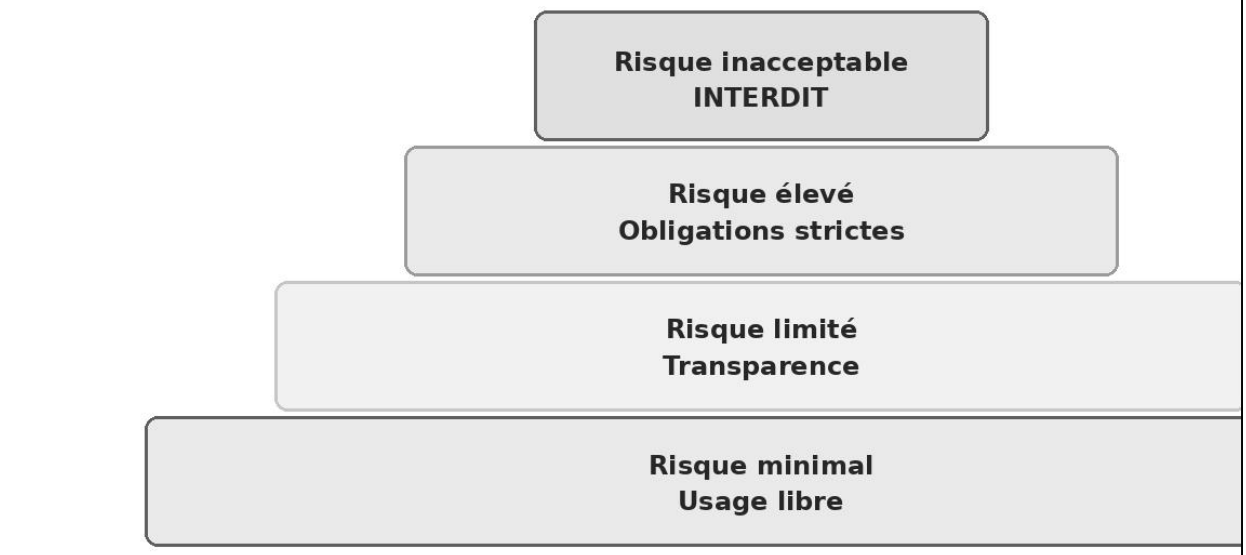
Cadre	Organisme	Année	Principes clés	Force juridique
Principes sur l'IA	**OCDE**	2019/2024	Inclusion, transparence, robustesse, sécurité, responsabilité	Volontaire
Recommandation sur l'éthique	**UNESCO**	2021	Droits humains, équité, durabilité, transparence, responsabilité	Volontaire
AI Act	Union européenne	2024	Approche par niveaux de risque, interdictions, obligations	Contraignant (UE)
Hiroshima AI Process	**G7**	2023	Transparence, sécurité, protection des utilisateurs	Volontaire
NIST AI RMF	**NIST (USA)**	2023	Gouvernance, cartographie, mesure, gestion des risques	Volontaire (USA)

Tableau 7.2 : Principaux cadres éthiques et réglementaires de l'IA. Sources : OCDE, UNESCO, Commission européenne, NIST.

7.3.1 L'AI Act européen : la régulation la plus avancée

L'AI Act européen, adopté en juin 2024, est le premier texte législatif contraignant au monde qui régule spécifiquement l'intelligence artificielle. Son approche repose sur une classification des systèmes d'IA par niveau de risque.

Figure 7.3 : La pyramide de risques de l'AI Act européen



L'AI Act est important pour plusieurs raisons qui dépassent les frontières européennes. D'abord, toute entreprise qui vend des systèmes d'IA sur le marché européen : qu'elle soit basée aux États-Unis, en Chine ou en Algérie : doit s'y conformer. C'est ce qu'on appelle l'effet Bruxelles. Ensuite, ce texte établit un vocabulaire et une méthodologie de classification du risque qui influenceront probablement d'autres régulations nationales, y compris futures en Algérie.

7.4 La notion de transparence et d'explicabilité (XAI)

L'explicabilité de l'IA (eXplainable AI, ou XAI) désigne l'ensemble des techniques qui permettent de comprendre et d'expliquer les décisions d'un modèle. Elle répond à une exigence à la fois technique (comment améliorer le modèle), éthique (pourquoi le modèle a-t-il pris cette décision), et juridique (droit à l'explication prévu par le RGPD).

Plusieurs approches existent. LIME (Local Interpretable Model-Agnostic Explanations) explique localement une prédiction en approximant le modèle complexe par un modèle simple autour de l'exemple à expliquer. SHAP (SHapley Additive exPlanations) attribue à chaque variable d'entrée une contribution à la prédiction, en s'inspirant de la théorie des jeux. Ces outils permettent par exemple de répondre à la question : « Pourquoi mon dossier de crédit a-t-il été refusé ? » avec une explication variable par variable, accessible à un non-spécialiste [1].

Références bibliographiques : Chapitre 7-Complément

Cadres éthiques et régulation internationale

Références complémentaires

[1] Lundberg, S. M. & Lee, S.-I. (2017). « A Unified Approach to Interpreting Model Predictions ». Advances in Neural Information Processing Systems, vol. 30. arXiv : 1705.07874. DOI : 10.48550/arXiv.1705.07874.

Chapitre 8 : Cybersécurité, automatisation intelligente et souveraineté numérique

Objectifs pédagogiques

- Comprendre les nouvelles menaces cybernétiques liées à l'IA.
- Saisir les mécanismes des attaques adversariales.
- Comprendre ce qu'est l'automatisation intelligente et ses effets sur le travail.
- Définir la souveraineté numérique et ses dimensions.
- Situer l'Algérie dans la géopolitique des données et de l'IA.

8.1 L'IA au service de la cybersécurité (et contre elle)

L'intelligence artificielle transforme radicalement le paysage de la cybersécurité. Elle est à la fois un outil de défense puissant et un amplificateur des capacités offensives des attaquants. Comprendre cette double nature est indispensable pour saisir les enjeux de souveraineté numérique.

8.1.1 L'IA défensive

Du côté défensif, l'IA permet de détecter des anomalies dans les comportements réseau beaucoup plus rapidement qu'un humain. Un système d'IA entraîné sur le comportement normal d'un réseau peut identifier en temps réel une intrusion ou une exfiltration de données : même si l'attaque emprunte un chemin jamais vu. Des outils comme Microsoft Defender, CrowdStrike Falcon ou IBM QRadar intègrent tous des composantes d'IA.

8.1.2 L'IA offensive : nouvelles armes, nouvelles menaces

Du côté offensif, l'IA démultiplie les capacités des attaquants. Trois menaces méritent une attention particulière.

Le phishing augmenté : l'IA générative permet de produire des milliers de courriels de hameçonnage personnalisés : adaptés au style d'écriture de l'expéditeur imité, à la situation professionnelle de la cible, à l'actualité du moment. Ces courriels sont beaucoup plus convaincants que les phishings génériques.

Les deepfakes des vidéos ou des enregistrements audio générés par IA imitent de manière convaincante le visage ou la voix d'une personnalité réelle. En 2024, une entreprise de Hong Kong a perdu 25 millions de dollars après qu'un employé a été trompé par un deepfake vidéo imitant son directeur financier lors d'une visioconférence.

Les attaques adversariales des modifications imperceptibles pour l'œil humain dans une image ou un texte peuvent tromper un modèle d'IA avec une très haute probabilité.

Exemple : Attaque adversariale

En ajoutant quelques pixels soigneusement choisis à l'image d'un panda (imperceptibles pour un humain), il est possible de faire croire à un classifieur d'images que le panda est en réalité un gibbon : avec une confiance de 99,3%. C'est une attaque adversariale [1]. Ces vulnérabilités ont des implications critiques pour les systèmes de reconnaissance faciale dans les aéroports ou les systèmes de conduite autonome.

8.2 L'automatisation intelligente

L'automatisation intelligente (ou RPA augmentée par l'IA) consiste à automatiser des tâches complexes qui combinent règles explicites et apprentissage automatique. À la différence de l'automatisation classique (qui ne fait que répéter), l'automatisation intelligente peut gérer des exceptions, s'adapter à des variations et améliorer ses performances au fil du temps.

8.2.1 Impact sur le marché du travail

Le rapport du Forum économique mondial de 2023 estime que 85 millions de postes pourraient être remplacés par l'automatisation d'ici 2030, mais que 97 millions de nouveaux postes adaptés à l'ère IA pourraient être créés [2]. Ces projections restent très incertaines, mais elles signalent une transformation, pas nécessairement une disparition du travail.

Ce qui est documenté avec plus de certitude, c'est la transformation des tâches. Les tâches routinières, qu'elles soient physiques ou cognitives, sont les plus exposées. Les tâches qui mobilisent le jugement, la créativité, la relation humaine et la négociation politique sont moins menacées, mais pas à l'abri de transformations importantes.

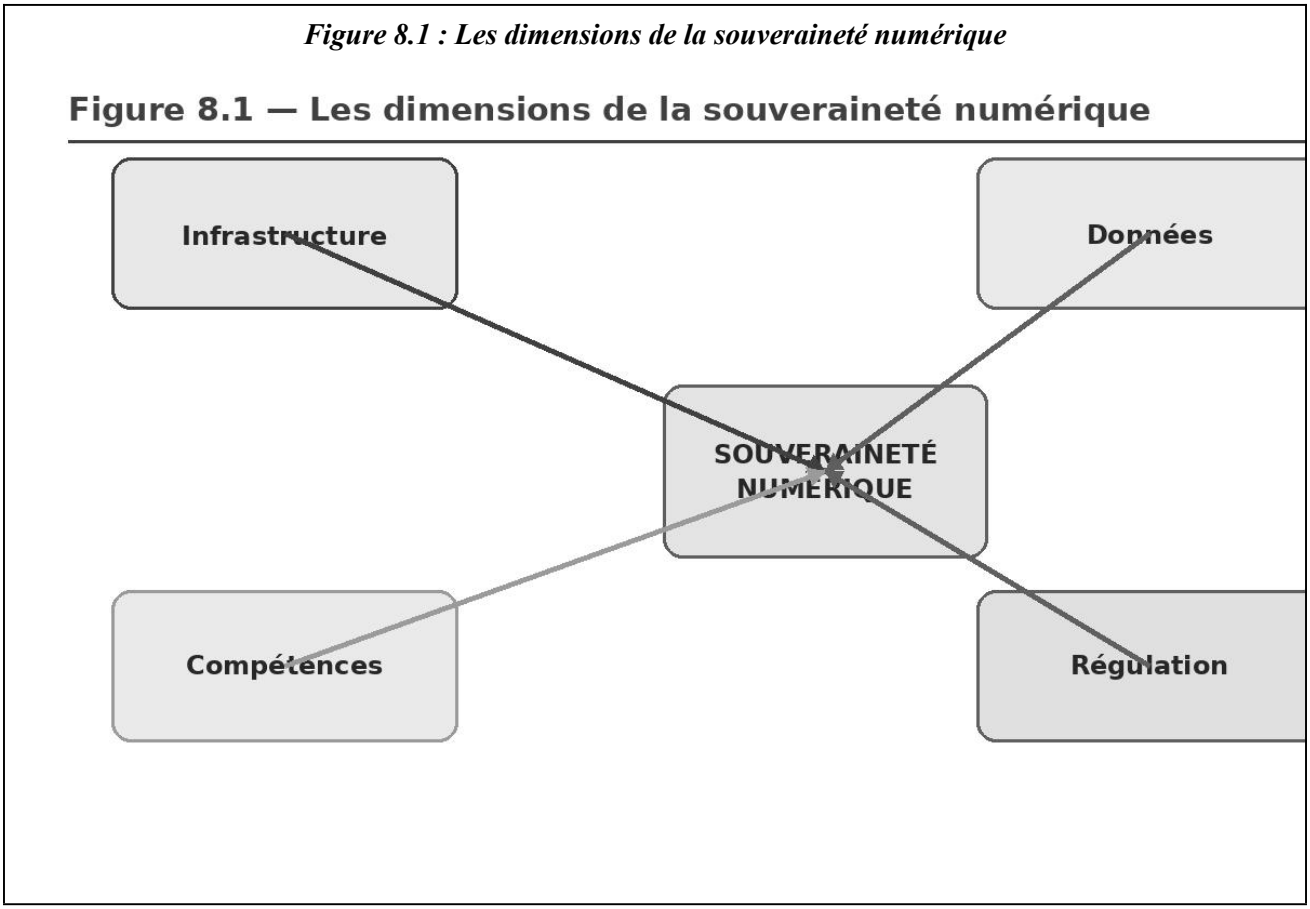
Type de tâche	Exposition à l'automatisation	Exemples
Routinière physique	Très élevée	Assemblage, manutention, saisie
Routinière cognitive	Élevée	Comptabilité basique, saisie de données, traitement de formulaires
Non routinière analytique	Moyenne	Analyse de données, recherche, rédaction standard
Non routinière interpersonnelle	Faible	Négociation, diplomatie, soin, enseignement
Non routinière créative	Faible à moyenne	Conception, innovation, direction artistique

Tableau 8.1 : Exposition des types de tâches à l'automatisation. Source : adapté de [2].

8.3 Souveraineté numérique de quoi s'agit-il ?

La souveraineté numérique désigne la capacité d'un État à contrôler les technologies numériques qu'il utilise, les données qu'il produit, et les normes qui régissent ces technologies sur son territoire [3]. C'est une notion multidimensionnelle qui touche à l'infrastructure, aux données, aux algorithmes et à la réglementation.

Figure 8.1 : Les dimensions de la souveraineté numérique



8.3.1 L'Algérie face aux enjeux de souveraineté numérique

L'Algérie dispose d'un cadre juridique initial avec la loi 18-07 du 10 juin 2018 relative à la protection des données personnelles, et la création de l'Autorité Nationale de Protection des Données Personnelles (ANPDP). Ces instruments constituent une base, mais plusieurs défis structurels subsistent : la dépendance aux infrastructures cloud étrangères (principalement Amazon AWS, Microsoft Azure et Google Cloud), la faiblesse des corpus de données en arabe et en darija disponibles pour entraîner des modèles locaux, et le manque de spécialistes formés à l'intersection de l'IA et des politiques publiques.

Le label « Projet Innovant » de startup.dz et la Stratégie nationale du numérique témoignent d'une prise de conscience. La construction d'une véritable autonomie technologique dans le domaine de l'IA exige cependant des investissements soutenus, des partenariats régionaux (avec d'autres pays arabes et africains) et une formation universitaire adaptée, qui est précisément ce à quoi ce module contribue.

À retenir sur le Chapitre 8

L'IA est à la fois un outil de cyberdéfense (détection d'anomalies) et un amplificateur des attaques (phishing ciblé, deepfakes, attaques adversariales).

L'automatisation intelligente transforme le travail : surtout les tâches routinières cognitives sans nécessairement supprimer les emplois.

La souveraineté numérique a quatre dimensions : infrastructure, données, algorithmes, normes.

L'Algérie dispose d'un cadre juridique initial (loi 18-07, ANPDP) mais reste dépendante de fournisseurs étrangers sur les dimensions infrastructurelles et algorithmiques.

Construire une souveraineté numérique réelle suppose des investissements, des partenariats et des formations spécialisées.

Références bibliographiques : Chapitre 8

Cybersécurité, automatisation intelligente et souveraineté numérique

Références citées

[1] Goodfellow, I. J. et al. (2014). « Explaining and Harnessing Adversarial Examples ». ICLR 2015. arXiv : 1412.6572. DOI : 10.48550/arXiv.1412.6572.

[2] World Economic Forum (2023). The Future of Jobs Report 2023. Geneva : WEF, 296 p. https://www.weforum.org/publications/the-future-of-jobs-report-2023/.

[3] Floridi, L. (2020). « The Fight for Digital Sovereignty : What It Is, and Why It Matters, Especially for the EU ». Philosophy & Technology, vol. 33, p. 369-378. DOI : 10.1007/s13347-020-00423-6.

AXE 3

APPLICATIONS EN SCIENCES POLITIQUES

تطبيقات في العلوم السياسية

Cet axe applique les bases techniques des deux premiers aux domaines qui définissent les sciences politiques. Trois chapitres examinent successivement la gouvernance et les services publics intelligents, la lutte contre la désinformation et la protection des processus démocratiques, et la géopolitique de l'IA dans les relations internationales. Chaque chapitre mobilise les notions techniques acquises pour éclairer des enjeux politiques concrets, avec des exemples documentés et des études de cas vérifiables.

Chapitre 9 : Intelligence artificielle, gouvernance et services publics

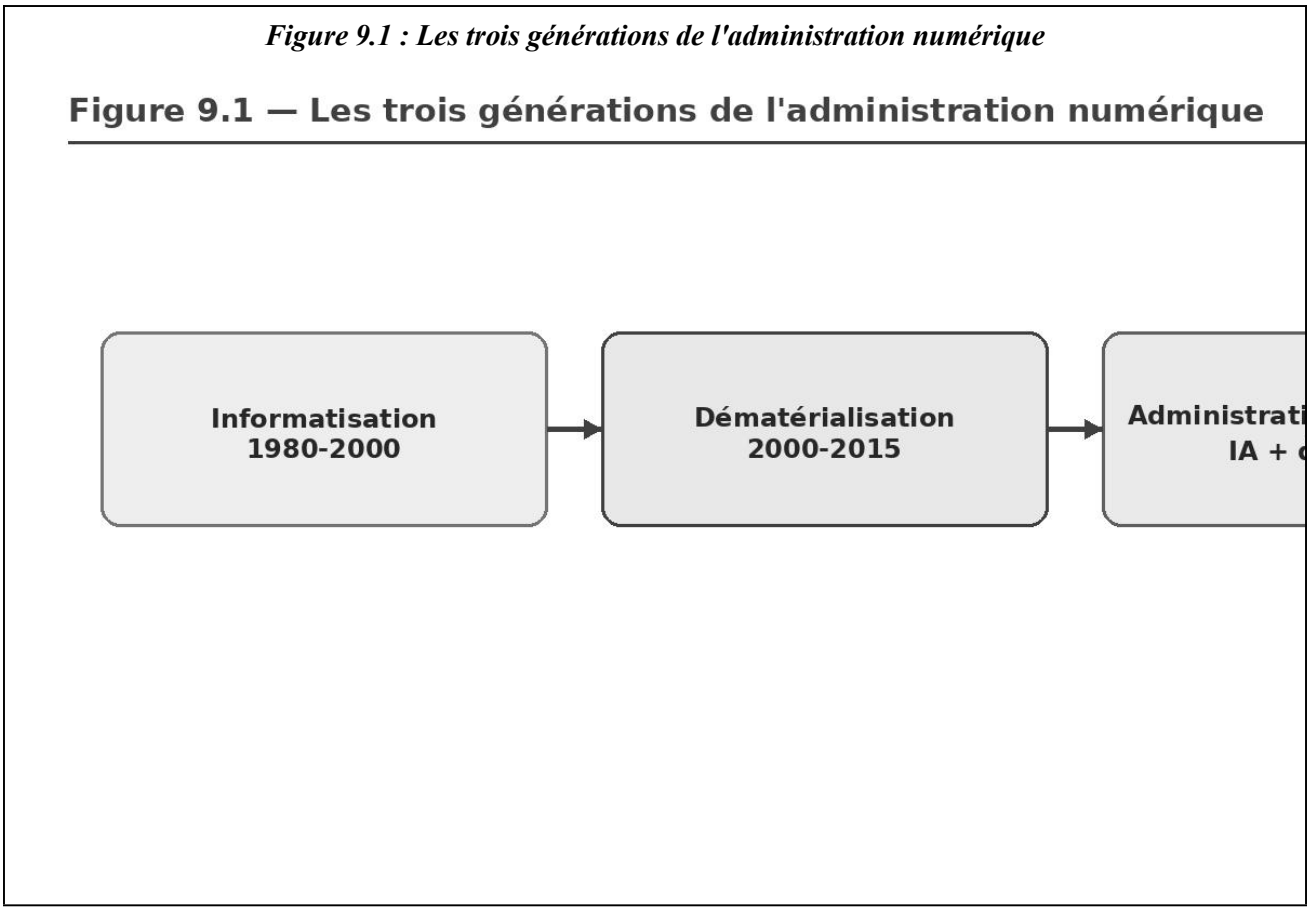
Objectifs pédagogiques

- Comprendre les trois générations de l'administration numérique.
- Identifier les domaines d'application prioritaires de l'IA dans les services publics.
- Analyser les conditions de réussite et les risques d'une IA publique légitime.
- Connaître les modèles internationaux de référence et leurs principes.
- Situer l'expérience algérienne dans ce panorama.

9.1 Du formulaire papier à l'administration algorithmique

L'administration publique a traversé en trente ans trois mutations numériques successives qui se superposent plutôt qu'elles ne se remplacent. La première génération, dans les années 1980-1990, consiste simplement à informatiser les tâches existantes : les formulaires restent, mais leur traitement est réalisé par des logiciels plutôt que par des employés qui tapent sur des machines à écrire. La deuxième génération, à partir des années 2000, dématérialise les démarches : l'usager peut les accomplir en ligne, sans se déplacer. La troisième génération, en cours depuis le milieu des années 2010, est l'algorithmisation : l'IA ne se contente plus de numériser une procédure existante, elle prend des décisions, classe des dossiers, prédit des risques, personnalise les services.

Figure 9.1 : Les trois générations de l'administration numérique



9.2 Applications sectorielles de l'IA dans les services publics

9.2.1 La santé publique

L'aide au diagnostic médical par IA est l'une des applications les plus matures. Des systèmes entraînés sur des millions de radiographies ou d'images de scanner atteignent des performances comparables à celles de spécialistes humains pour détecter certains cancers ou pathologies rétiniennes [1]. En Algérie, l'intégration de tels outils dans les hôpitaux universitaires représente une opportunité réelle face à la pénurie de certains spécialistes dans les régions éloignées.

La gestion des épidémies est un autre domaine. Lors de la pandémie de COVID-19, des modèles d'IA ont été utilisés pour prédire la propagation géographique, optimiser la distribution de vaccins et analyser les effets des mesures de confinement. Leurs résultats ont été inégaux, soulignant que la qualité des prévisions dépend directement de la qualité des données épidémiologiques disponibles.

9.2.2 L'éducation

Les systèmes de tutorat intelligent (Intelligent Tutoring Systems, ITS) adaptent le rythme et le contenu de l'enseignement au niveau de chaque élève [2]. Les plateformes d'orientation universitaire (comme le système algérien de pré-inscription des bacheliers) utilisent des algorithmes pour gérer des centaines de milliers de candidatures simultanément. Ces systèmes soulèvent des questions d'équité : si l'algorithme optimise un critère de mérite scolaire, il peut reproduire et amplifier les inégalités d'accès à l'éducation de qualité.

9.2.3 La justice et la sécurité

Les usages de l'IA dans la justice sont parmi les plus controversés. La prédiction de risque de récidive (comme COMPAS, étudié au chapitre 7) est un exemple de ce que le droit appelle parfois la « justice prédictive ». Des systèmes de reconnaissance faciale sont utilisés dans les aéroports et aux frontières pour l'identification. Des algorithmes analysent les données de délinquance pour orienter les patrouilles de police : un usage dont le caractère discriminatoire a été documenté [3].

9.2.4 La fiscalité et la détection de fraude

La détection automatique de fraude fiscale est l'un des usages les plus efficaces et les mieux acceptés de l'IA publique. Les systèmes apprennent les patterns des déclarations frauduleuses avérées, et signalent automatiquement les dossiers suspects pour révision humaine. L'efficacité est documentée : le Service général des impôts français rapporte une augmentation significative des redressements ciblés grâce à ces outils, avec une réduction du contrôle aléatoire injustifié.

9.3 Conditions de réussite et risques

9.3.1 Les conditions de réussite

Condition	Description	Exemple
Infrastructure de qualité	Réseau haut débit, datacenters fiables, identité numérique	L'Estonie a construit X-Road avant de déployer l'e-gouvernement
Qualité des données	Données complètes, récentes, représentatives et protégées	Un système de santé mal renseigné produit des recommandations dangereuses
Transparence algorithmique	L'usager peut comprendre et contester la décision	Obligation de motiver les décisions automatisées (RGPD, AI Act)
Recours humain	Toute décision automatisée peut être réexaminée par un humain	Droit au réexamen humain inscrit dans l'AI Act européen
Formation des agents	Les fonctionnaires comprennent les outils qu'ils utilisent	Programme national de formation (comme l'initiative Data Literacy en

Condition	Description	Exemple
		France)

Tableau 9.1 : Conditions de réussite d'une IA publique légitime.

9.3.2 Le risque d'exclusion numérique

Numériser un service sans alternative ne le rend pas plus accessible : cela peut l'exclure des personnes non connectées, mal voyantes, peu scolarisées ou résidant dans des zones à mauvaise connexion. La dématérialisation accélérée des services publics avait créé des situations d'abandon administratif pour les publics les plus vulnérables. Toute politique d'IA publique doit préserver des canaux alternatifs.

À retenir sur le Chapitre 9

L'administration numérique traverse trois générations : informatisation, dématérialisation, algorithmisation.

Les applications prioritaires sont la santé, l'éducation, la justice et la fiscalité : chacune avec des promesses et des risques spécifiques.

Une IA publique légitime exige : infrastructure, qualité des données, transparence, recours humain et formation.

La numérisation sans alternative risque d'exclure les publics les plus vulnérables.

Le modèle estonien repose sur l'identité numérique, l'interopérabilité et le principe 'once only'.

Activité 9.1 : Analyse d'un service public numérique

Choisissez un service public en ligne algérien (portail des téléservices, plateforme de pré-inscription universitaire, Téléchifa, etc.). Analysez-le selon les cinq critères du tableau 9.1. Identifiez ses points forts, ses lacunes et formulez trois recommandations d'amélioration en mobilisant les notions techniques des chapitres 1 à 6.

Références bibliographiques : Chapitre 9

Intelligence artificielle, gouvernance et services publics

Références citées

[1] Esteva, A. et al. (2017). « Dermatologist-level classification of skin cancer with deep neural networks ». Nature, vol. 542, n° 7639, p. 115-118. DOI : 10.1038/nature21056.

[2] VanLehn, K. (2011). « The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems ». Educational Psychologist, vol. 46, n° 4, p. 197-221. DOI : 10.1080/00461520.2011.611369.

[3] Lum, K. & Isaac, W. (2016). « To predict and serve? ». Significance, vol. 13, n° 5, p. 14-19. DOI : 10.1111/j.1740-9713.2016.00960.x.

Extension Chapitre 9 : Modèles internationaux comparés

9.4 Quatre modèles de gouvernement numérique

Au-delà du modèle estonien, trois autres expériences méritent d'être présentées pour leur intérêt comparatif.

9.4.1 Le modèle singapourien

Singapour a déployé depuis 2014 son programme Smart Nation, visant à intégrer les technologies dans tous les secteurs de la ville-État. L'application SingPass regroupe plus de mille services publics accessibles avec une seule identité numérique. Le système LifeSG organise les démarches selon les étapes de la vie : naissance, scolarité, emploi, retraite. Les résultats sont impressionnants en termes d'efficacité, mais le modèle singapourien bénéficie de conditions particulières : population réduite (6 millions), gouvernement fort et stable, taux d'urbanisation de 100%, revenus élevés. Ces conditions ne sont pas généralisables telles quelles.

9.4.2 Le modèle des Émirats arabes unis

Les Émirats ont créé en 2017 le premier ministère mondial de l'IA. L'application UAE PASS centralise les démarches administratives avec une identité numérique nationale. La stratégie nationale d'IA 2031 vise à faire des Émirats un hub mondial. Ce qui est particulièrement pertinent pour l'Algérie est l'investissement dans les modèles de langage arabes (Falcon, Jais) et la politique d'attraction de talents internationaux. Les ressources financières pétrolières jouent ici un rôle que l'Algérie peut en partie reproduire.

9.4.3 Le modèle rwandais

Le Rwanda offre une expérience particulièrement instructive pour les pays africains. Malgré un revenu par habitant faible, le gouvernement rwandais a déployé des téléservices ambitieux, utilisé des drones pour la livraison de médicaments dans les zones rurales (partenariat avec Zipline), et intégré l'IA dans le suivi épidémique. La clé du succès rwandais n'est pas la richesse : c'est la cohérence de la vision politique et la priorité donnée à l'infrastructure numérique de base.

9.4.4 Comparaison et leçons pour l'Algérie

Pays	Indice EGDI ONU 2024	Identité numérique	Points transposables en Algérie
Estonie	0,95 (très élevé)	e-ID universelle (2002)	Interopérabilité X-Road, principe once-only, cybersécurité

Pays	Indice EGDI ONU 2024	Identité numérique	Points transposables en Algérie
Singapour	0,93 (très élevé)	SingPass (2003)	Intégration des services par étapes de vie, gouvernance
Émirats arabes unis	0,92 (très élevé)	**UAE PASS (2020)**	Modèles de langage arabes, attraction de talents, vision stratégique
Rwanda	Intermédiaire en hausse	En développement	Cohérence politique, drones médicaux, priorité infrastructure
Algérie	0,57 (intermédiaire)	CNI biométrique	À développer : interopérabilité, identité numérique unifiée, cloud souverain

Tableau 9.2 : Comparaison internationale des modèles d'e-gouvernement. Source : ONU E-Government Survey 2024.

9.5 La détection automatique de fraude : un cas d'usage détaillé

La détection de fraude est l'une des applications d'IA les plus matures et les mieux acceptées dans l'administration publique. Détaillons son fonctionnement pour un service fiscal.

Un système de détection de fraude fiscale utilise typiquement un algorithme de forêt aléatoire ou un réseau de neurones entraîné sur des milliers de dossiers de fraudes avérées et de déclarations légitimes. Il apprend à identifier des patterns suspects : incohérence entre chiffre d'affaires déclaré et dépenses observées, variation anormale entre deux exercices, structures de transactions inhabituelles.

Le système produit pour chaque dossier un score de risque entre 0 et 1. Les dossiers au-dessus d'un certain seuil sont signalés aux inspecteurs humains pour révision. La décision finale reste humaine. L'algorithme optimise la sélection des dossiers à contrôler, ce qui multiplie l'efficacité des équipes d'inspection sans les remplacer.

Points de vigilance pour l'usage administratif

1. Biais géographique : si les contrôles passés ont ciblé certaines régions ou secteurs, le modèle amplifiera ce biais.
2. Dérive du modèle : les schémas de fraude évoluent. Un modèle entraîné en 2020 peut être obsolète en 2025 si l'environnement fiscal a changé.
3. Transparence : le contribuable dont le dossier est sélectionné a-t-il le droit de savoir que c'est un algorithme qui l'a ciblé ?
4. Recours : le droit à un examen humain de la décision doit être garanti, conformément aux exigences de l'AI Act (article 14) et du RGPD (article 22).

Chapitre 10 : Intelligence artificielle, désinformation et processus démocratiques

Objectifs pédagogiques

- Distinguer désinformation, mésinformation et malinformation avec précision.
- Comprendre techniquement comment l'IA produit des deepfakes et du contenu synthétique.
- Identifier les mécanismes d'amplification algorithmique de la désinformation.
- Connaître les approches techniques et politiques de détection et de lutte.
- Analyser les risques spécifiques pour les processus électoraux et démocratiques.

10.1 Cartographier l'écosystème de la désinformation

Avant toute analyse, les définitions s'imposent. La désinformation désigne la production et la diffusion intentionnelles de fausses informations dans le but de tromper. La mésinformation désigne la diffusion non intentionnelle d'informations incorrectes, par méconnaissance sincère. La malinformation désigne la diffusion d'informations vraies dans un contexte ou avec une intention malveillante. Ces trois catégories ont des implications juridiques et politiques différentes [1].

10.1.1 L'IA générative comme accélérateur

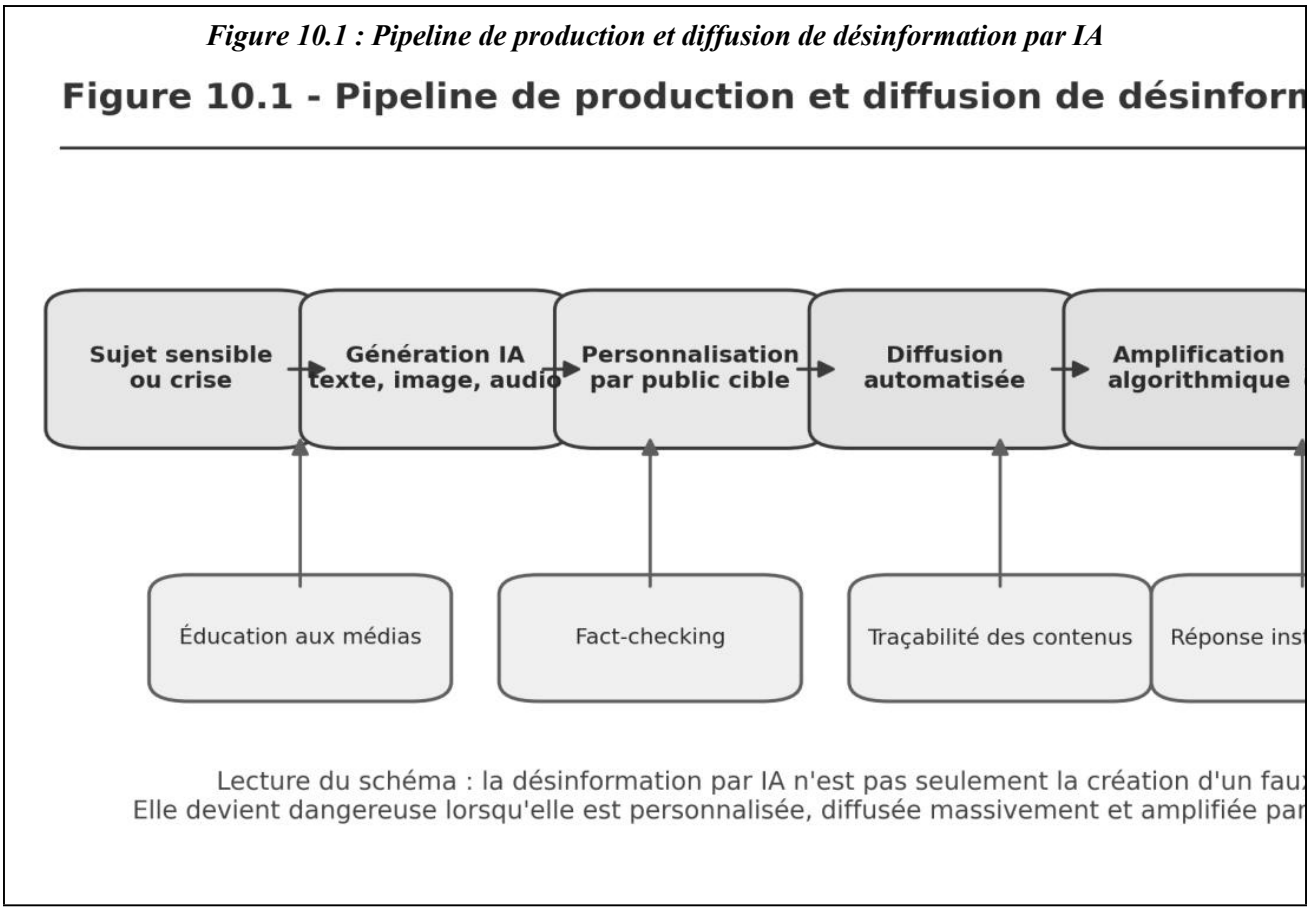
Jusqu'en 2022, produire de la désinformation de qualité : un faux article convaincant, une image truquée réaliste, une vidéo deepfake : exigeait des ressources humaines et techniques significatives. L'IA générative a radicalement modifié cette équation. Trois ruptures techniques sont responsables de cette transformation.

Première rupture : le coût de production effondré. Un faux article de presse convaincant peut être généré en quelques secondes par un LLM, en dizaines de langues simultanément, à un coût quasi nul. Une opération d'influence qui nécessitait autrefois une équipe de rédacteurs peut être automatisée.

Deuxième rupture : la qualité des deepfakes. Les modèles de diffusion (Stable Diffusion, DALL-E, Midjourney) génèrent des images indiscernables des photographies réelles pour la majorité des observateurs non entraînés. Les modèles de clonage vocal (ElevenLabs, etc.) reproduisent la voix d'une personne à partir de quelques secondes d'enregistrement.

Troisième rupture : la personnalisation à grande échelle. L'IA permet d'adapter le message désinformant au profil psychologique de chaque cible : âge, convictions politiques, peurs documentées, ce qui en démultiplie l'efficacité.

Figure 10.1 : Pipeline de production et diffusion de désinformation par IA



10.2 Les deepfakes : mécanismes et détection

10.2.1 Comment fonctionne un deepfake vidéo

Un deepfake vidéo est produit par un réseau antagoniste génératif (GAN) ou, plus récemment, par un modèle de diffusion. Dans l'approche GAN, deux réseaux s'affrontent : le générateur produit de fausses images du visage cible; le discriminateur essaie de distinguer le vrai du faux. Au fil de l'entraînement, le générateur devient de plus en plus convaincant. Le résultat est une vidéo dans laquelle le visage d'une personne est substitué à celui d'une autre de manière quasi imperceptible.

10.2.2 La détection technique

Plusieurs approches techniques permettent de détecter les contenus synthétiques. Les classifieurs entraînés sur des données authentiques et falsifiées apprennent à reconnaître les artefacts caractéristiques des modèles génératifs. Le watermarking numérique intègre des marques imperceptibles dans les contenus générés pour permettre leur identification ultérieure. La coalition C2PA (Coalition for Content Provenance and Authenticity), réunissant Adobe, Microsoft, BBC et d'autres acteurs, développe des standards de traçabilité de la provenance des contenus [2].

Mais la course attaquant-défenseur reste structurellement déséquilibrée : les outils de génération progressent aussi vite, sinon plus vite, que les outils de détection. C'est pourquoi les réponses purement techniques sont insuffisantes.

10.3 IA et processus électoraux

10.3.1 Le micro-ciblage électoral

Le micro-ciblage électoral consiste à segmenter l'électorat et à personnaliser les messages politiques selon le profil de chaque segment. L'affaire Cambridge Analytica (2018) en a révélé les mécanismes des données Facebook obtenues sans consentement explicite ont permis de construire des profils psychographiques de millions d'électeurs américains et britanniques, auxquels des messages publicitaires politiques soigneusement adaptés ont été adressés [3]. L'IA générative amplifie maintenant cette capacité en permettant de produire des messages personnalisés à grande échelle et à coût marginal quasi nul.

10.3.2 L'analyse automatisée du discours politique

Les outils de traitement automatique du langage (NLP) permettent d'analyser des corpus massifs de discours politiques pour détecter des thèmes, mesurer des sentiments, cartographier l'évolution de la rhétorique d'un acteur. Ces techniques, utilisées en recherche académique depuis les années 2000, sont maintenant à la portée de toute organisation disposant d'une connexion Internet. Elles ouvrent des possibilités nouvelles pour la surveillance des discours de haine, le suivi des campagnes de désinformation, et l'analyse comparative des programmes électoraux.

10.3.3 Les réponses politiques et réglementaires

Instrument	Portée	Disposition principale
AI Act européen (2024)	Union européenne	Obligation de signaler les contenus générés par IA dans les publicités politiques
Digital Services Act (2022)	Union européenne	Obligations de transparence des algorithmes de recommandation pour les très grandes plateformes
Loi anti-manipulation (France, 2018)	France	Permet la saisine du juge en période électorale pour retrait de fausses informations
Coalition C2PA	International (volontaire)	Standard de traçabilité de la provenance des contenus numériques
Loi 18-07 (Algérie, 2018)	Algérie	Protection des données personnelles : applicable aux données électorales collectées en ligne

Tableau 10.1 : Principaux instruments réglementaires face à la désinformation par IA.

À retenir sur le Chapitre 10

Désinformation (intentionnelle), mésinformation (non intentionnelle) et malinformation (vraie mais malveillante) sont trois phénomènes distincts.

L'IA générative a effondré les coûts de production de contenus trompeurs et permis leur personnalisation à grande échelle.

Les deepfakes utilisent des GAN ou des modèles de diffusion pour substituer ou générer des visages et des voix.

La détection technique (classifieurs, watermarking, C2PA) est nécessaire mais insuffisante face à la rapidité des avancées génératives.

Le micro-ciblage électoral combiné à l'IA générative constitue un risque structurel pour l'intégrité des processus démocratiques.

Activité 10.1 : Fact-checking assisté par outils numériques

À partir d'une image ou d'une vidéo politique récente circulant sur les réseaux sociaux, effectuez une vérification en plusieurs étapes : (a) recherche d'image inversée (Google Images, TinEye) pour vérifier la provenance; (b) vérification de la date de première publication (InVID/WeVerify); (c) consultation d'une organisation de fact-checking (AFP Factuel, Fatabyyano, Misbar); (d) si possible, utilisation d'un détecteur de deepfake disponible en ligne (Deepware, Sensity). Rédigez un rapport d'une page concluant sur l'authenticité du contenu et la méthode utilisée.

Références bibliographiques : Chapitre 10

Intelligence artificielle, désinformation et processus démocratiques

Références citées

[1] Wardle, C. & Derakhshan, H. (2017). Information Disorder : Toward an interdisciplinary framework for research and policy making. Conseil de l'Europe, DGI(2017)09. https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77.

[2] C2PA (Coalition for Content Provenance and Authenticity). Technical Specification v2.1. https://c2pa.org/specifications/specifications/2.1/index.html.

[3] Cadwalladr, C. & Graham-Harrison, E. (2018). « Revealed : 50 million Facebook profiles harvested for Cambridge Analytica ». The Guardian, 17 mars 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Extension Chapitre 10 : Analyse computationnelle des discours politiques

10.4 Les outils d'analyse du discours politique

Le traitement automatique du langage ouvre des possibilités nouvelles pour l'analyse des discours politiques. Trois techniques méritent une présentation détaillée.

10.4.1 L'analyse de sentiments

L'analyse de sentiments (sentiment analysis) classe automatiquement un texte selon son ton émotionnel : positif, négatif, ou neutre. Appliquée à des milliers de tweets ou de commentaires sur une décision gouvernementale, elle permet de mesurer en temps réel la réaction de l'opinion publique. Des outils commerciaux comme Brandwatch ou des bibliothèques open source comme VADER ou TextBlob (en anglais) permettent cette analyse.

Les limites sont importantes. L'ironie et le second degré trompent les classifieurs. Les expressions idiomatiques et les références culturelles locales sont souvent mal interprétées. En arabe et encore plus en darija, les performances sont nettement inférieures à l'anglais. Une réaction qui semble « positive » dans les mots peut être sarcastique dans le contexte culturel : une nuance que l'algorithme rate souvent.

10.4.2 La détection de thèmes (topic modeling)

Le topic modeling (modélisation thématique) est une technique d'apprentissage non supervisé qui identifie automatiquement les thèmes dominants dans un corpus de textes. L'algorithme le plus utilisé, LDA (Latent Dirichlet Allocation), suppose que chaque document est un mélange de plusieurs thèmes, chaque thème étant caractérisé par une distribution de mots.

Un chercheur qui analyse dix années de discours présidentiels peut utiliser le topic modeling pour identifier comment les thèmes prioritaires évoluent dans le temps sans lire manuellement des milliers de pages. L'algorithme ne nomme pas les thèmes : c'est le chercheur qui les interprète à partir des mots les plus représentatifs. Cette nécessité d'interprétation humaine est un garde-fou important contre une automatisation aveugle.

10.4.3 La détection de la rhétorique populiste

Des chercheurs en science politique computationnelle ont développé des classifieurs capables d'identifier les traits rhétoriques du discours populiste dans des textes : opposition peuple/élites, appel au leader charismatique, discours anti-institutions [1]. Ces outils permettent d'analyser systématiquement des corpus de discours politiques pour mesurer le degré de populisme de différents acteurs.

Ces méthodes ont été appliquées notamment aux discours de partis européens (projet PopuList) et latino-américains. Leur application aux discours arabes reste limitée faute de données annotées suffisantes : un déficit de recherche que les universités algériennes pourraient contribuer à combler.

10.5 Carte des acteurs de la vérification des faits dans le monde arabe

Organisation	Pays	Langue principale	Site
Fatabyyano	Jordanie (régional)	Arabe standard	https://fatabyyano.net
Misbar	Émirats arabes unis	Arabe + anglais	https://misbar.com
AFP Factuel (Maghreb)	France/régional	Arabe + français	https://factuel.afp.com
Kashif (Algérie)	Algérie	Arabe + darija (émergent)	En développement
CheckDesk / Meedan	International	Multilingue dont arabe	https://meedan.com

Tableau 10.2 : Principaux acteurs du fact-checking dans le monde arabe.

Références bibliographiques : Chapitre 10-Extension

Analyse computationnelle des discours politiques

Références

[1] Rooduijn, M. & Pauwels, T. (2011). « Measuring Populism : Comparing Two Methods of Content Analysis ». West European Politics, vol. 34, n° 6, p. 1272-1283. DOI : 10.1080/01402382.2011.616665.

Chapitre 11 : Intelligence artificielle, relations internationales et géopolitique

Objectifs pédagogiques

- Comprendre les dimensions géopolitiques de la compétition autour de l'IA.
- Identifier les usages militaires et diplomatiques de l'IA.
- Analyser la position de l'Algérie et du monde arabe dans cette compétition.
- Saisir les enjeux des négociations internationales sur la régulation de l'IA.
- Évaluer les risques et opportunités pour les pays en développement.

11.1 La géopolitique de l'IA : une nouvelle course aux armements

La course mondiale à l'intelligence artificielle ressemble, par certains aspects, aux grandes rivalités technologiques du XXe siècle : la course à l'espace, la course au nucléaire. Comme elles, elle oppose des blocs aux intérêts divergents, mobilise des ressources colossales, et est perçue comme un enjeu de sécurité nationale.

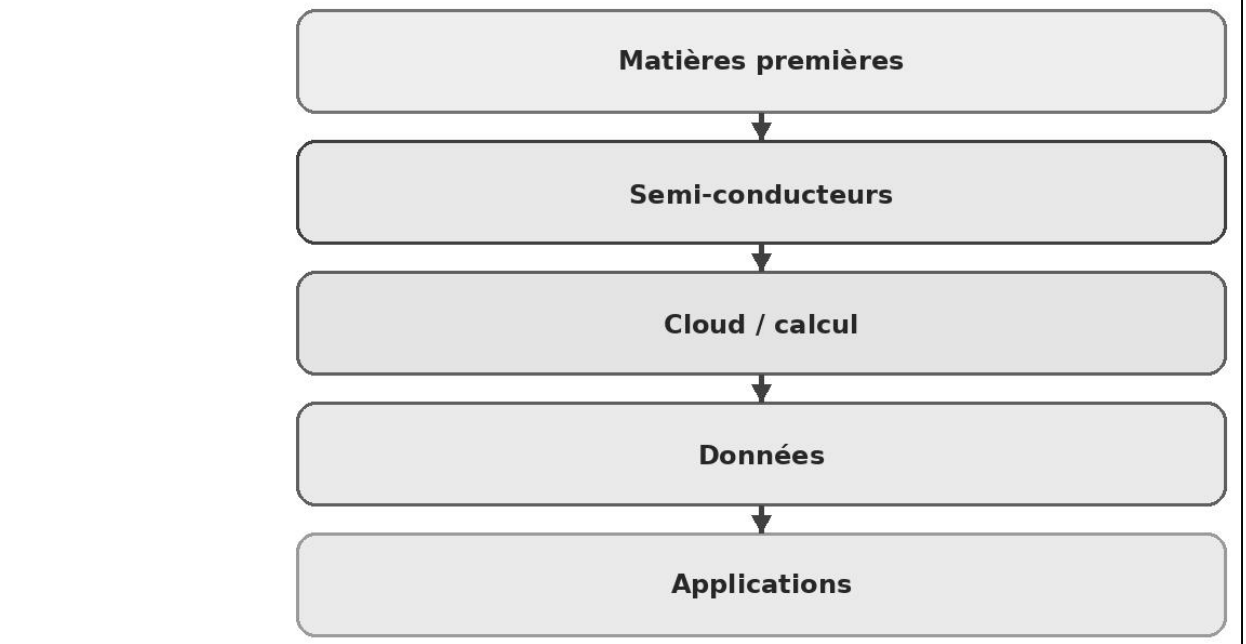
Mais elle s'en distingue aussi profondément. L'IA n'est pas une technologie uniquement militaire : elle traverse tous les secteurs économiques et sociaux. Elle ne nécessite pas d'installations gigantesques et identifiables : un groupe de chercheurs avec les bonnes données et les bons GPU peut produire des systèmes très avancés. Et elle dépend d'une chaîne d'approvisionnement mondiale extrêmement concentrée, dont le point de passage le plus étroit est la fabrication de semi-conducteurs avancés.

11.1.1 Les semi-conducteurs : le point de passage le plus étroit

Toute IA moderne repose sur des puces électroniques et les plus performantes sont fabriquées par une poignée d'entreprises. TSMC (Taiwan Semiconductor Manufacturing Company) produit la majorité des puces les plus avancées du monde. ASML (Pays-Bas) est le seul fabricant mondial des machines de lithographie extrême ultraviolet (EUV) nécessaires pour graver les circuits les plus fins. Ces deux goulots d'étranglement sont au cœur de la rivalité sino-américaine [1].

Depuis 2022, les États-Unis ont imposé des restrictions croissantes à l'exportation de puces et d'équipements de fabrication vers la Chine, avec l'objectif explicite de ralentir le développement des capacités militaires chinoises en IA. En réponse, la Chine a accéléré ses investissements dans la production nationale de semi-conducteurs. Ces tensions ont des effets concrets sur tous les pays qui souhaitent développer leur capacité en IA, y compris les pays du Sud, qui se retrouvent dans une situation de dépendance technologique structurelle.

Figure 11.1 : La chaîne de valeur mondiale de l'IA



11.2 L'IA dans la défense et la sécurité internationale

11.2.1 Les systèmes d'armes autonomes

Les systèmes d'armes létaux autonomes (SALA), parfois appelés « robots tueurs », désignent des dispositifs militaires capables de sélectionner et d'engager des cibles sans intervention humaine directe. Leur développement actif par plusieurs puissances militaires : États-Unis, Chine, Russie, Israël : soulève des questions juridiques et éthiques sans précédent, qui est responsable d'un crime de guerre commis par un système autonome ? Comment garantir le respect du droit international humanitaire si la décision de tuer est prise par un algorithme ?

Le débat international se déroule dans le cadre de la Convention sur certaines armes classiques (CCAC) à Genève. Plusieurs États, dont beaucoup de pays du Sud, plaident pour un traité d'interdiction. Les grandes puissances technologiques militaires résistent à toute régulation contraignante. L'Algérie participe à ces négociations en tant que membre de l'ONU et de l'Union africaine.

11.2.2 La cyber-guerre augmentée par l'IA

L'IA transforme la cyber-guerre en rendant les attaques plus rapides, plus adaptatives et plus difficiles à attribuer. Des logiciels malveillants qui s'adaptent automatiquement aux défenses rencontrées, des campagnes de phishing personnalisées à grande échelle, des attaques sur les infrastructures critiques (énergie, eau, finance) coordonnées par des systèmes automatiques : ces capacités étaient réservées aux États les plus sophistiqués il y a dix ans, elles se diffusent maintenant à des acteurs non étatiques disposant de ressources moyennes.

11.3 La diplomatie de l'IA

11.3.1 Les négociations internationales sur la régulation

La gouvernance mondiale de l'IA est un champ diplomatique en formation. Plusieurs enceintes multilatérales jouent un rôle.

Enceinte	Instrument principal	Position
Union européenne	AI Act (2024)	Régulation contraignante basée sur les risques : approche la plus avancée
ONU / UNESCO	Recommandation éthique (2021), Pacte numérique (2024)	Cadre de principes non contraignant, universellement négocié
OCDE	Principes IA (2019, révisés 2024)	Principes volontaires pour les pays membres
G7 / G20	Hiroshima AI Process (2023)	Coordination entre grandes économies, sans contrainte juridique
Nations du Sud	Coalitions à l'ONU, Union africaine	Plaident pour un accès équitable et une régulation tenant compte du développement

Tableau 11.1 : Principaux cadres de gouvernance internationale de l'IA.

11.3.2 La position de l'Algérie et du monde arabe

Les pays arabes présentent une situation très hétérogène face à l'IA. Les Émirats arabes unis ont fait de l'IA un axe stratégique national, avec un ministère dédié (créé en 2017), des investissements massifs dans les modèles de langage arabes (Falcon, Jais) et une ambition de diversification post-pétrolière. L'Arabie saoudite investit des dizaines de milliards dans le secteur à travers son programme Vision 2030. Ces pays bénéficient de ressources financières considérables et de la capacité d'attirer des talents internationaux.

L'Algérie, avec un capital humain réel de nombreux ingénieurs et docteurs en informatique, souvent formés à l'étranger et une population de plus de 45 millions d'habitants représentant un marché significatif, dispose d'atouts que des politiques publiques cohérentes pourraient valoriser. Les priorités identifiées dans la littérature spécialisée sur les pays en développement convergent : constituer des corpus de données en langues locales, former des spécialistes à l'interface de la technique et des politiques publiques, développer des partenariats régionaux plutôt que de tenter de concurrencer seul les géants américains et chinois [2].

À retenir sur le Chapitre 11

La compétition mondiale pour l'IA se concentre sur trois ressources : les semi-conducteurs, les données et les talents.

TSMC et ASML sont les goulots d'étranglement industriels de la chaîne mondiale : d'où l'enjeu stratégique de Taïwan.

Les systèmes d'armes autonomes soulèvent des questions juridiques non résolues sur la responsabilité dans les conflits armés.

La gouvernance internationale de l'IA se construit dans plusieurs enceintes (ONU, OCDE, G7, AI Act) avec des approches divergentes.

Les pays du Sud, dont l'Algérie, ont intérêt à des partenariats régionaux plutôt qu'à une compétition frontale avec les grands blocs technologiques.

Activité 11.1 : Analyse géopolitique

À partir des notions du chapitre et du tableau 11.1, rédigez une note de deux pages sur la position que devrait adopter l'Algérie dans les négociations onusiennes sur la régulation de l'IA. Précisez : (a) quels principes défendre en priorité (équité d'accès, souveraineté des données, interdiction des armes autonomes, etc.); (b) avec quels pays former des coalitions; (c) quels sont les risques d'une absence de position nationale claire.

Références bibliographiques : Chapitre 11

Intelligence artificielle, relations internationales et géopolitique

Références citées

[1] Miller, C. (2022). Chip War : The Fight for the World's Most Critical Technology. New York : Scribner, 464 p. ISBN 978-1-982172-03-5.

[2] Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M. & Floridi, L. (2018). « Artificial Intelligence and the 'Good Society' : the US, EU, and UK approach ». Science and Engineering Ethics, vol. 24, n° 2, p. 505-528. DOI : 10.1007/s11948-017-9901-7.

Extension Chapitre 11 : L'IA dans la diplomatie et l'intelligence stratégique

11.4 L'IA au service de la diplomatie

11.4.1 La veille diplomatique automatisée

Les ministères des affaires étrangères et les services de renseignement utilisent des systèmes d'IA pour surveiller en temps réel des milliers de sources d'information : presse internationale, réseaux sociaux, communiqués officiels, dépêches d'agences. Le volume d'information disponible dépasse largement les capacités d'analyse humaine. Un système bien calibré peut détecter en quelques heures l'émergence d'une crise, identifier les acteurs impliqués et croiser les informations pour évaluer leur fiabilité.

Des projets académiques comme GDELT (Global Database of Events, Language, and Tone) indexent des millions d'articles de presse du monde entier chaque jour et permettent de visualiser l'intensité des événements géopolitiques en temps réel [1]. Ces ressources, partiellement accessibles au public, illustrent les capacités des systèmes professionnels utilisés par les grandes chancelleries.

11.4.2 La traduction diplomatique automatique

La traduction automatique a fait des progrès spectaculaires depuis l'adoption des modèles Transformer. Pour les diplomates, cela signifie un accès immédiat à des documents officiels dans des langues rares ou une pre-lecture rapide d'une déclaration étrangère avant la traduction humaine officielle. Mais les nuances diplomatiques : la différence entre « préoccupé » et « alarmé », entre « espère » et « exige » : restent l'apanage du traducteur humain expert.

La traduction de l'arabe vers le français et l'anglais pose des défis spécifiques liés à la richesse morphologique et à la polysémie. L'Algérie, dont la diplomatie s'exprime souvent en arabe, en français et en anglais, dispose d'un intérêt stratégique direct dans le développement de modèles de traduction haute qualité arabophones.

11.4.3 La prévision et la prévention des conflits

Des systèmes d'alerte précoce utilisant le ML analysent des indicateurs politiques, économiques et sociaux pour détecter les signaux précurseurs de crises violentes. Le projet ACLED (Armed Conflict Location and Event Data) recense quotidiennement les événements violents dans le monde et alimente des modèles prédictifs [2]. Ces outils sont utilisés par l'ONU, l'Union africaine et plusieurs ONG pour orienter les ressources de prévention vers les zones les plus à risque.

Leurs performances restent cependant limitées : les grandes ruptures politiques : révolutions, coups d'État : sont structurellement difficiles à prévoir, car elles résultent de dynamiques non linéaires et de décisions individuelles imprévisibles. Ces outils sont plus utiles pour suivre des crises en cours que pour les anticiper de manière fiable.

11.5 Enjeux pour les pays en développement : entre opportunité et dépendance

La question de fond pour des pays comme l'Algérie n'est pas « faut-il adopter l'IA ? » : la technologie se diffuse indépendamment des choix nationaux. La question est « comment l'adopter de manière à renforcer plutôt qu'à affaiblir l'autonomie nationale ? »

Trois stratégies ne s'excluent pas et se complètent. La stratégie d'adaptation consiste à utiliser et adapter des modèles étrangers existants : les modèles open source comme Llama ou Falcon peuvent être fine-tunés sur des données locales à coût raisonnable. La stratégie de spécialisation consiste à développer une expertise nationale dans des niches où les besoins locaux sont spécifiques : traitement de la darija, reconnaissance de l'écriture arabe manuscrite, modélisation des marchés agricoles locaux. La stratégie de coalition consiste à coopérer avec d'autres pays arabes et africains pour mutualiser les ressources : constitution de corpus communs, partage d'infrastructures, formation conjointe.

À retenir sur l' Extension Chapitre 11

La veille diplomatique automatisée démultiplie les capacités d'analyse sans remplacer le jugement politique humain.

La traduction automatique accélère l'accès à l'information étrangère mais reste limitée sur les nuances diplomatiques.

Les modèles de prévision des conflits sont utiles pour le suivi mais peu fiables pour l'anticipation des ruptures.

Pour les pays en développement, trois stratégies complémentaires : adaptation, spécialisation, coalition régionale.

Références bibliographiques : Chapitre 11-Extension

L'IA dans la diplomatie et l'intelligence stratégique

Références

[1] GDELT Project. Base de données mondiale d'événements géopolitiques indexés par IA. https://www.gdeltproject.org.

[2] ACLED : Armed Conflict Location and Event Data Project. https://acleddata.com.

Fiches récapitulatives : Résumés comparatifs

Vue d'ensemble comparative des onze chapitres

Cette fiche récapitulative synthétise les concepts centraux de chaque chapitre sous forme tabulaire pour faciliter la révision.

Chapitre	Notion centrale	Distinction clé	Application politique emblématique
Ch. 1 : IA : définition et histoire	L'IA apprend à partir de données	IA étroite (réelle) ≠ IA générale (hypothétique)	Classification des usages gouvernementaux par type d'IA
Ch. 2 : Données et algorithmes	Qualité des données = qualité du modèle	Algorithme classique ≠ algorithme apprenant	Données électorales et Big Data pour le ciblage
Ch. 3 : Machine Learning	Apprendre par l'erreur = ajuster les paramètres	Supervisé / non supervisé / par renforcement	Détection de fraude fiscale et scoring social
Ch. 4 : Réseaux de neurones	Représentation hiérarchique des données	CNN (images) ≠ Transformer (texte)	Modèles arabes : Falcon, Jais pour services publics
Ch. 5 : NLP et IA générative	LLM = prédiction du token suivant	Discriminatif (classe) ≠ génératif (crée)	Chatbots de campagne, traduction diplomatique
Ch. 6 : Vision, recommandation	CNN pour images, algorithmes de filtrage collaboratif	Bulle de filtre = polarisation par optimisation	Reconnaissance faciale aux frontières, fil d'actualité
Ch. 7 : Erreurs et biais	Hallucination ≠ mensonge intentionnel	Biais des données ≠ biais de conception	Affaire COMPAS : discrimination raciale automatisée
Ch. 8 : Cybersécurité	Deepfakes + phishing IA = nouvelles armes hybrides	Souveraineté infra ≠ souveraineté données ≠ normes	Loi 18-07 algérienne et ANPDP
Ch. 9 : Gouvernance	3 générations : informatisation → dématérialisation → algo	Conditions de légitimité (5 critères)	Modèle estonien vs pratiques algériennes
Ch. 10 : Désinformation	IA générative = coût de production effondré	Désinformation (intentionnelle) ≠ mésinformation	Cambridge Analytica et deepfake Zelensky
Ch. 11 : Géopolitique	Semi-conducteurs = goulot d'étranglement mondial	Chaîne de valeur (5 niveaux) : matières premières → applis	Position Algérie dans négociations ONU sur l'IA

Tableau récapitulatif : Synthèse des onze chapitres du polycopié.

Questions de révision transversales

Questions de compréhension

1. Expliquez en cinq phrases le fonctionnement d'un réseau de neurones profond. Incluez les notions de paramètre, de rétropropagation et de couche cachée.
2. Quelle est la différence entre l'apprentissage supervisé et l'apprentissage par renforcement ? Donnez un exemple politique pour chacun.
3. Comment un LLM génère-t-il son texte ? Pourquoi produit-il parfois des hallucinations ?
4. Qu'est-ce qu'un biais algorithmique ? Citez ses trois sources principales avec un exemple pour chacune.
5. Expliquez ce qu'est un deepfake et comment il est techniquement produit.

Questions d'analyse politique

6. Analysez l'affaire COMPAS en mobilisant les notions de biais algorithmique, surapprentissage et métriques d'équité.
7. Quelles sont les conditions de réussite d'une IA publique légitime ? Illustrez avec le modèle estonien.
8. En quoi l'IA générative constitue-t-elle une rupture pour la désinformation politique ?
9. Quels sont les goulots d'étranglement de la chaîne mondiale de l'IA ? Quelles en sont les implications pour l'Algérie ?
10. Comparez les approches européenne (AI Act) et américaine de la régulation de l'IA. Quelle approche conviendrait mieux au contexte algérien ?

CONCLUSION GÉNÉRALE

Ce que vous emportez et ce qui reste à faire

Onze chapitres, trois axes, un même fil conducteur : l'intelligence artificielle est une technique au service de décisions, et toute décision technique est, en dernière analyse, un choix politique. Revenir sur les acquis essentiels permet de consolider ce fil.

Ce que vous savez maintenant

Vous savez que l'IA n'est pas de la magie mais une famille de fonctions mathématiques qui apprennent à partir de données. Vous savez distinguer l'apprentissage supervisé de l'apprentissage non supervisé, le Machine Learning du Deep Learning, un réseau de neurones classique d'un Transformer. Vous comprenez comment un modèle de langage génère son texte mot après mot, et pourquoi cette mécanique produit inévitablement des hallucinations. Vous avez compris comment un système de recommandation optimise l'engagement plutôt que la vérité, et comment ce choix d'optimisation a des effets politiques documentés.

Vous connaissez les biais algorithmiques et leurs trois sources : les données, la conception, le déploiement. Vous savez que corriger un biais peut en révéler un autre, et que les définitions formelles de l'équité sont parfois mathématiquement incompatibles. Vous comprenez les enjeux de cybersécurité liés à l'IA : les deepfakes, les attaques adversariales, le phishing augmenté et les dimensions de la souveraineté numérique qui se jouent au-delà de la simple protection des données personnelles.

Dans le domaine de la gouvernance, vous avez vu comment trois générations d'administration numérique se superposent, quelles conditions permettent à une IA publique d'être légitime, et pourquoi la numérisation sans alternative peut créer de l'exclusion. Dans le domaine de la désinformation, vous avez compris comment l'IA générative a effondré les coûts de production du faux contenu et transformé structurellement la menace pour les processus démocratiques. Dans le domaine des relations internationales, vous avez identifié les goulots d'étranglement de la chaîne mondiale de l'IA, les enjeux des négociations réglementaires et la position spécifique des pays du Sud.

Ce qui reste à construire

Ce polycopié couvre les fondements. L'intelligence artificielle est un domaine en mouvement si rapide que des pans entiers de ce qui s'écrit aujourd'hui sera dépassé dans dix-huit mois. Ce qui ne changera pas, ce sont les questions à poser face à n'importe quel système d'IA, qui l'a construit ? Avec quelles données ? Pour optimiser quoi ? Qui en bénéficie ? Qui en supporte les risques ? Qui peut le contester ? Ces questions sont celles du politiste. Ce polycopié vous a donné les outils techniques pour y répondre avec précision.

La suite appartient à votre propre travail de veille, de lecture et de pratique. L'IA INDEX du Stanford HAI publie chaque année un état des lieux rigoureux et gratuit. Les rapports de l'OCDE AI Policy Observatory cartographient les politiques nationales. Les publications de l'IEEE, de l'ACM et d'arXiv rendent accessibles les avancées de la recherche. Et les plateformes comme ChatGPT, Claude ou Gemini sont elles-mêmes des objets d'étude que vous pouvez interroger, tester et critiquer à condition de ne jamais oublier ce que vous savez maintenant de leur fonctionnement.

Dix compétences acquises à l'issue de ce module

1. Définir l'IA avec précision et distinguer ses niveaux de capacités.
2. Expliquer le rôle des données dans la qualité et l'équité d'un modèle.
3. Distinguer les trois familles d'apprentissage et leurs applications.
4. Comprendre le fonctionnement d'un LLM et identifier ses hallucinations.
5. Analyser techniquement un système de recommandation et ses effets politiques.
6. Identifier les sources et conséquences d'un biais algorithmique.
7. Évaluer les risques cyber liés à l'IA (deepfakes, phishing, adversarial).
8. Analyser une politique d'IA publique selon les critères de légitimité.
9. Distinguer les instruments réglementaires (AI Act, DSA, loi 18-07) et leurs portées.
10. Situer un pays dans la géopolitique mondiale de l'IA et identifier ses leviers d'action.

GLOSSAIRE TRILINGUE

Français : العربية : English

Ce glossaire regroupe les 42 termes essentiels abordés dans ce polycopié. Chaque entrée inclut le terme en français, son équivalent en arabe (standard), son équivalent en anglais, une définition courte et le numéro du chapitre où la notion est développée. Les termes sont classés alphabétiquement en français.

Français	العربية	English	Définition et chapitre
Agent conversationnel (chatbot)	روبوت المحادثة	Chatbot	Programme qui simule une conversation humaine, basé sur des règles ou sur un LLM. Ch. 6.
Algorithme	خوارزمية	Algorithm	Suite finie d'instructions précises permettant de résoudre un problème. Ch. 2.
Apprentissage automatique	تعلم آلي	Machine Learning (ML)	Famille de techniques permettant à un système d'apprendre à partir de données sans règles explicites. Ch. 3.
Apprentissage par renforcement	تعلم بالتعزيز	Reinforcement Learning (RL)	Apprentissage par essai-erreur avec récompenses et pénalités dans un environnement. Ch. 2.
Apprentissage profond	تعلم عميق	Deep Learning (DL)	ML utilisant des réseaux de neurones à plusieurs couches pour apprendre des représentations hiérarchiques. Ch. 3.
Apprentissage supervisé	تعلم موجَّه	Supervised Learning	Apprentissage à partir de données étiquetées (exemples avec la bonne réponse fournie). Ch. 2.
Apprentissage non supervisé	تعلم غير موجَّه	Unsupervised Learning	Apprentissage à partir de données non étiquetées : découverte de structures cachées. Ch. 2.
Architecture Transformer	بنية المحوِّل	Transformer	Architecture de réseau de neurones basée sur l'attention, à la base de tous les LLM modernes. Ch. 4.
Attaque adversariale	هجوم عدائي	Adversarial Attack	Modification imperceptible d'une entrée pour tromper un modèle d'IA avec haute confiance. Ch. 8.
Biais algorithmique	تحيز خوارزمي	Algorithmic Bias	Distorsion systématique dans les prédictions d'un modèle, défavorisant

Français	العربية	English	Définition et chapitre
			certains groupes. Ch. 7.
Big Data	بيانات ضخمة	Big Data	Données caractérisées par 5V : Volume, Vélocité, Variété, Véracité, Valeur. Ch. 2.
Bulle de filtre	فقاعة المرشح	Filter Bubble	Phénomène d'isolement informationnel créé par les algorithmes de recommandation personnalisés. Ch. 6.
Chatbot	روبوت المحادثة	Chatbot	Voir agent conversationnel. Ch. 6.
CNN (réseau convolutif)	شبكة التفاف عصبية	Convolutional Neural Network	Architecture de réseau de neurones spécialisée dans le traitement des images. Ch. 4.
Couche cachée	طبقة خفية	Hidden Layer	Couche intermédiaire d'un réseau de neurones entre l'entrée et la sortie. Ch. 4.
Deepfake	تزييف عميق	Deepfake	Contenu audiovisuel synthétique généré par IA imitant une personne réelle de façon trompeuse. Ch. 10.
Descente de gradient	الانحدار التدرجي	Gradient Descent	Algorithme d'optimisation qui ajuste les paramètres pour minimiser l'erreur du modèle. Ch. 3.
Désinformation	تضليل	Disinformation	Diffusion intentionnelle de fausses informations dans un but trompeur. Ch. 10.
Embedding	تمثيل متجهي	Embedding	Représentation numérique d'un mot ou d'une phrase sous forme de vecteur. Ch. 5.
Fine-tuning	ضبط دقيق	Fine-tuning	Adaptation d'un modèle de fondation pré-entraîné à une tâche spécifique. Ch. 4.
GAN	شبكة توليدية تعاكسية	Generative Adversarial Network	Architecture de deux réseaux concurrents (générateur et discriminateur) pour générer des contenus. Ch. 4.
Hallucination	هلوسة	Hallucination	Production par un LLM d'une information fausse présentée avec assurance. Ch. 7.
IA étroite (ANI)	ذكاء اصطناعي ضيق	Narrow AI (ANI)	Système IA performant sur une tâche précise uniquement : seul type réellement existant. Ch. 1.
IA générale (AGI)	ذكاء اصطناعي عام	Artificial General Intelligence	Système IA hypothétique capable de toutes les tâches cognitives humaines. N'existe pas. Ch. 1.

Français	العربية	English	Définition et chapitre
IA générative	ذكاء اصطناعي توليدي	Generative AI	IA capable de créer du contenu nouveau (texte, image, son, vidéo). Ch. 5.
LLM	نموذج لغوي كبير	Large Language Model	Modèle de langage pré-entraîné sur des corpus massifs, capable de générer du texte. Ch. 5.
Mésinformation	معلومات مغلوطة	Misinformation	Diffusion non intentionnelle d'informations incorrectes par croyance sincère. Ch. 10.
Micro-ciblage	استهداف دقيق	Microtargeting	Personnalisation des messages politiques selon le profil psychographique individuel. Ch. 10.
Modèle de diffusion	نموذج انتشار	Diffusion Model	Architecture générative qui crée des images en dé-bruisant progressivement un bruit aléatoire. Ch. 5.
Modèle de fondation	نموذج أساسي	Foundation Model	Grand modèle pré-entraîné sur données massives, adaptable à de nombreuses tâches. Ch. 4.
NLP / TAL	معالجة اللغة الطبيعية	Natural Language Processing	Branche de l'IA traitant du langage humain (texte et parole). Ch. 5.
Neurone artificiel	خلية عصبية اصطناعية	Artificial Neuron	Unité de calcul de base d'un réseau de neurones : somme pondérée + fonction d'activation. Ch. 4.
Overfitting	إفراط في التدريب	Overfitting (surapprentissage)	Mémorisation des données d'entraînement au détriment de la généralisation. Ch. 3.
Paramètre	معلمة	Parameter	Valeur numérique ajustée lors de l'entraînement pour minimiser l'erreur du modèle. Ch. 3.
Pipeline	مسار المعالجة	Pipeline	Séquence d'étapes de traitement des données, de la collecte à l'entraînement du modèle. Ch. 2.
RAG	توليد معزز باسترجاع	Retrieval-Augmented Generation	Technique combinant un LLM avec une recherche dans une base externe pour améliorer la précision. Ch. 6.
Rétropropagation	الانتشار العكسي	Backpropagation	Algorithme d'apprentissage qui remonte l'erreur de la sortie vers l'entrée pour ajuster les poids. Ch. 4.
Semi-conducteur	شبه موصل	Semiconductor	Composant électronique (puce) sur lequel repose toute l'infrastructure de l'IA

Français	العربية	English	Définition et chapitre
			moderne. Ch. 11.
Souveraineté numérique	السيادة الرقمية	Digital Sovereignty	Capacité d'un État à contrôler ses infrastructures, données, algorithmes et normes numériques. Ch. 8.
Superintelligence (ASI)	ذكاء فائق	Artificial Superintelligence	IA hypothétique dépassant l'humain dans tous les domaines cognitifs. Prospective. Ch. 1.
Token	وحدة نصية	Token	Unité de base du texte traitée par un LLM (mot ou sous-mot). Ch. 5.
Watermarking	علامة مائية رقمية	Digital Watermarking	Marque imperceptible intégrée dans un contenu pour permettre d'en identifier la provenance. Ch. 10.

BIBLIOGRAPHIE GÉNÉRALE

I. Ouvrages fondamentaux sur l'intelligence artificielle

Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. Cambridge (Mass.) : MIT Press, 800 p. ISBN 978-0-262-03561-3. Disponible gratuitement : https://www.deeplearningbook.org.

Russell, S. & Norvig, P. (2021). Artificial Intelligence : A Modern Approach, 4e éd. Hoboken : Pearson, 1136 p. ISBN 978-0-13-461099-3. [accès payant]

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. New York : Springer, 738 p. ISBN 978-0-387-31073-2. Disponible gratuitement sur le site officiel de Microsoft Research.

Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3e éd. Sebastopol : O'Reilly Media, 851 p. ISBN 978-1-098-12597-4. [accès payant]

Mitchell, M. (2019). Artificial Intelligence : A Guide for Thinking Humans. New York : Farrar, Straus and Giroux, 336 p. ISBN 978-0-374-25783-5. [accès payant]

II. Articles scientifiques fondateurs

Turing, A. M. (1950). « Computing Machinery and Intelligence ». Mind, vol. LIX, n° 236, p. 433-460. DOI : 10.1093/mind/LIX.236.433.

McCulloch, W. S. & Pitts, W. (1943). « A logical calculus of the ideas immanent in nervous activity ». The Bulletin of Mathematical Biophysics, vol. 5, n° 4, p. 115-133. DOI : 10.1007/BF02478259.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). « Learning representations by back-propagating errors ». Nature, vol. 323, n° 6088, p. 533-536. DOI : 10.1038/323533a0.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). « Deep learning ». Nature, vol. 521, n° 7553, p. 436-444. DOI : 10.1038/nature14539.

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, vol. 30, p. 5998-6008. arXiv : 1706.03762. DOI : 10.48550/arXiv.1706.03762.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017). « ImageNet classification with deep convolutional neural networks ». Communications of the ACM, vol. 60, n° 6, p. 84-90. DOI : 10.1145/3065386.

Silver, D. et al. (2016). « Mastering the game of Go with deep neural networks and tree search ». Nature, vol. 529, n° 7587, p. 484-489. DOI : 10.1038/nature16961.

Bommasani, R. et al. (2021). « On the Opportunities and Risks of Foundation Models ». arXiv : 2108.07258. DOI : 10.48550/arXiv.2108.07258.

Strubell, E., Ganesh, A. & McCallum, A. (2019). « Energy and Policy Considerations for Deep Learning in NLP ». Proceedings of the 57th Annual Meeting of the ACL, p. 3645-3650. DOI : 10.18653/v1/P19-1355.

Ji, Z. et al. (2023). « Survey of Hallucination in Natural Language Generation ». ACM Computing Surveys, vol. 55, n° 12, art. 248. DOI : 10.1145/3571730.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). « Distributed Representations of Words and Phrases and their Compositionality ». Advances in NeurIPS, vol. 26. arXiv : 1310.4546. DOI : 10.48550/arXiv.1310.4546.

Goodfellow, I. J. et al. (2014). « Explaining and Harnessing Adversarial Examples ». ICLR 2015. arXiv : 1412.6572. DOI : 10.48550/arXiv.1412.6572.

Antoun, W., Baly, F. & Hajj, H. (2020). « AraBERT : Transformer-based Model for Arabic Language Understanding ». Proceedings of the 4th Workshop on Open-Source Arabic Corpora, LREC 2020. arXiv : 2003.00104. DOI : 10.48550/arXiv.2003.00104.

III. Biais, équité et éthique

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. (2021). « A Survey on Bias and Fairness in Machine Learning ». ACM Computing Surveys, vol. 54, n° 6, art. 115. DOI : 10.1145/3457607.

Buolamwini, J. & Gebru, T. (2018). « Gender Shades : Intersectional Accuracy Disparities in Commercial Gender Classification ». Proceedings of Machine Learning Research, vol. 81, p. 77-91. https://proceedings.mlr.press/v81/buolamwini18a.html.

Bellamy, R. K. E. et al. (2019). « AI Fairness 360 : An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias ». IBM Journal of Research and Development, vol. 63, n° 4/5, art. 4:1. DOI : 10.1147/JRD.2019.2942287.

O'Neil, C. (2016). Weapons of Math Destruction. New York : Crown Publishers, 272 p. ISBN 978-0-553-41881-1. [accès payant : traduction française disponible]

Pasquale, F. (2015). The Black Box Society. Cambridge (Mass.) : Harvard University Press, 320 p. ISBN 978-0-674-36827-9. [accès payant]

IV. Gouvernance, droit et régulation

OECD (2024). Recommendation of the Council on Artificial Intelligence (révisée 2024). OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

UNESCO (2022). Recommandation sur l'éthique de l'intelligence artificielle. Paris : UNESCO, 44 p. https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre.

Parlement européen et Conseil (2024). Règlement (UE) 2024/1689 : AI Act. Journal officiel de l'UE, L 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689.

Parlement européen et Conseil (2022). Règlement (UE) 2022/2065 : Digital Services Act (DSA). Journal officiel de l'UE, L 277, p. 1-102. https://eur-lex.europa.eu/eli/reg/2022/2065.

Algérie. Loi n° 18-07 du 10 juin 2018 relative à la protection des données à caractère personnel. Journal officiel de la République algérienne démocratique et populaire, n° 34 du 10 juin 2018. https://www.joradp.dz.

Floridi, L. (2020). « The Fight for Digital Sovereignty : What It Is, and Why It Matters ». Philosophy & Technology, vol. 33, p. 369-378. DOI : 10.1007/s13347-020-00423-6.

V. Désinformation, médias et processus démocratiques

Wardle, C. & Derakhshan, H. (2017). Information Disorder : Toward an interdisciplinary framework for research and policy making. Conseil de l'Europe, DGI(2017)09. https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77.

Vosoughi, S., Roy, D. & Aral, S. (2018). « The spread of true and false news online ». Science, vol. 359, n° 6380, p. 1146-1151. DOI : 10.1126/science.aap9559.

Pariser, E. (2011). The Filter Bubble : What the Internet Is Hiding from You. New York : Penguin Press, 294 p. ISBN 978-1-59420-300-8. [accès payant]

Huszár, F. et al. (2022). « Algorithmic amplification of politics on Twitter ». Proceedings of the National Academy of Sciences, vol. 119, n° 1, e2025334119. DOI : 10.1073/pnas.2025334119.

Chesney, R. & Citron, D. (2019). « Deep Fakes : A Looming Challenge for Privacy, Democracy, and National Security ». California Law Review, vol. 107, p. 1753-1820. DOI : 10.15779/Z38RV0D15J.

VI. Géopolitique de l'IA et rapports institutionnels

Miller, C. (2022). Chip War : The Fight for the World's Most Critical Technology. New York : Scribner, 464 p. ISBN 978-1-982172-03-5. [accès payant]

Stanford HAI (2025). Artificial Intelligence Index Report 2025. Stanford : Human-Centered Artificial Intelligence. https://aiindex.stanford.edu/report/.

World Economic Forum (2023). The Future of Jobs Report 2023. Geneva : WEF, 296 p. https://www.weforum.org/publications/the-future-of-jobs-report-2023/.

Bommasani, R. et al. (2021). « On the Opportunities and Risks of Foundation Models ». arXiv : 2108.07258. DOI : 10.48550/arXiv.2108.07258.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M. & Floridi, L. (2018). « Artificial Intelligence and the 'Good Society' : the US, EU, and UK approach ». Science and Engineering Ethics, vol. 24, n° 2, p. 505-528. DOI : 10.1007/s11948-017-9901-7.

OECD AI Policy Observatory. Politiques nationales d'IA et base de données interactive. https://oecd.ai.

Lum, K. & Isaac, W. (2016). « To predict and serve? ». Significance, vol. 13, n° 5, p. 14-19. DOI : 10.1111/j.1740-9713.2016.00960.x.

VanLehn, K. (2011). « The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems ». Educational Psychologist, vol. 46, n° 4, p. 197-221. DOI : 10.1080/00461520.2011.611369.

VII. Ressources en ligne vérifiées et mises à jour

arXiv.org. Dépôt de preprints en informatique et IA. https://arxiv.org/list/cs.AI/recent : Consulté régulièrement pour les avancées récentes.

Papers With Code. Classements des meilleurs modèles avec code source disponible. https://paperswithcode.com.

Hugging Face. Bibliothèque de modèles open source et jeux de données. https://huggingface.co.

ANPDP (Algérie). Autorité Nationale de Protection des Données Personnelles. https://www.anpdp.dz.

NIST AI Risk Management Framework. National Institute of Standards and Technology. https://airc.nist.gov/Home.

C2PA. Coalition for Content Provenance and Authenticity : standards de traçabilité. https://c2pa.org.

e-Estonia. Solutions d'e-gouvernement estoniennes : référence mondiale. https://e-estonia.com.

Fatabyyano. Plateforme de fact-checking en langue arabe. https://fatabyyano.net.

IEEE Transactions on Neural Networks and Learning Systems. Revue scientifique de référence. DOI préfixe : 10.1109/TNNLS. [accès payant via bibliothèques universitaires]

FIN DU POLYCOPIÉ

L'Intelligence Artificielle

Dr. BOUNOUNI Mahdi

Faculté de Droit et des Sciences Politiques

Sciences Politiques