

Référentiel

Les compétences numériques en intelligence artificielle

Développer les compétences numériques
essentielles pour maîtriser l'IA



Avant-propos

Pix crée de nouveaux contenus d'évaluation et de formation sur des thématiques fortes et d'actualité. Après la cybersécurité et le numérique responsable, c'est maintenant **l'intelligence artificielle** qui est priorisée.

L'intelligence artificielle (IA) transforme notre société à une vitesse sans précédent, touchant tous les aspects de notre vie quotidienne, de la santé à l'éducation, en passant par l'économie et les services publics. Face à cette révolution technologique, il est crucial de garantir que ces avancées bénéficient à tous, de manière équitable et responsable et ne soient pas réservées à une élite technologique ou économique. Un référentiel accessible à tous permet de démocratiser l'accès aux connaissances et aux compétences nécessaires pour comprendre et utiliser l'IA, réduisant ainsi les inégalités numériques. La transparence dans le développement et l'utilisation de l'IA est fondamentale pour instaurer la confiance du public. Un référentiel clair et compréhensible permet à chacun de comprendre les enjeux éthiques, les limites et les potentialités de l'IA.



Opération soutenue par l'État dans le cadre du dispositif « Démonstration en conditions réelles, amorçage et premières commerciales » de France 2030, opéré par la Banque des Territoires (Caisse des dépôts)



BANQUE des
TERRITOIRES



Caisse
des Dépôts

Contexte

Pix conçoit des référentiels de compétences numériques. Ils décrivent des savoirs et savoir-faire qui ont vocation à une certaine stabilité et à être indépendants des évolutions de surface des outils, ou des dispositifs numériques.

Ces référentiels servent de cadre pour la création de contenus d'évaluation et de contenus de formation.

Pour concevoir des contenus d'évaluation, les compétences sont décomposées en « acquis », unités de savoirs et savoir-faire observables qu'il est possible d'évaluer. Pour concevoir des contenus de formation, les compétences sont traduites en « objectifs pédagogiques », qui guident la conception de modules de formation.

Le présent référentiel cadre porte sur l'intelligence artificielle. Il devrait permettre aux partenaires de Pix d'intégrer les offres d'évaluation et de formations dans leurs propres formations. Il est également destiné aux apprenants pour leur permettre d'identifier les notions importantes à connaître dans cette thématique.

Inspirations

L'**UNESCO** a publié, en 2024, deux référentiels* : l'un destiné à un public d'étudiants et l'autre à un public d'enseignants. Ils se découpent en 4 domaines sur trois niveaux de progression inspirés de la taxonomie de Bloom. Ces référentiels mettent en avant la préservation de l'agentivité humaine et l'engagement de la responsabilité humaine dans la prise de décision dans toute tâche outillée ou augmentée par un système piloté par l'IA.

Ce référentiel ne consacre pas un domaine à l'agentivité, considérant qu'elle doit être prise en compte transversalement à tous les domaines décrits.

Référentiel de compétences en IA pour les apprenants

Aspects des compétences	Niveaux de progression		
	Comprendre	Appliquer	Créer
Perspective centrée sur l'humain	Agentivité humaine	Responsabilité humaine	Citoyenneté à l'ère de l'IA
Éthique de l'IA	Intériorisation de l'éthique (« <i>Embodied ethics</i> »)	Utilisation sûre et responsable	Éthique dès la conception (« <i>Ethics by design</i> »)
Techniques et applications de l'IA	Fondements de l'IA	Compétences pour l'application	Création d'outils d'IA
Conception de systèmes d'IA	Délimitation des problèmes	Conception de l'architecture	Itérations et boucles de rétroaction

Source : (Miao, F., et al., 2024)

* [Miao, F., Shiohira, K., Lao, N., Référentiel de compétences en IA pour les apprenants, UNESCO, 2024.](#)

Le DigComp 2.2** comporte une annexe titrée « **interaction citoyenne avec des systèmes d'IA** » (p77-83). Elle présente une liste de 73 exemples de compétences en intelligence artificielle. Cette liste est organisée en 4 thèmes :

1. Que font et ne font pas les systèmes d'IA ?
2. Comment fonctionnent les systèmes d'IA ?
3. Quand interagir avec des systèmes d'IA ?
4. Les défis et les enjeux éthique de l'IA
5. Attitudes en matière d'agentivité et de contrôle humains

La structure du présent référentiel s'inspire largement de cette découpe en **5 axes**.

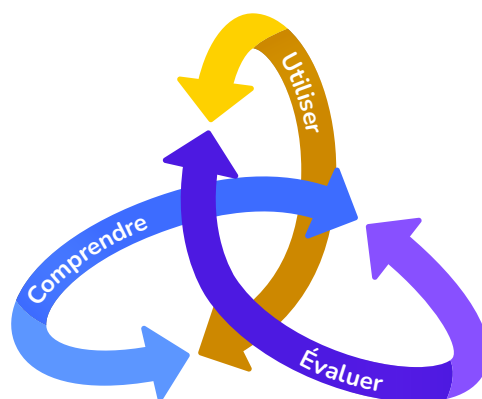
Le DigComp s'appuie toujours sur le référentiel des compétences numériques en cinq domaines et préconise d'introduire les compétences en IA dans chacune des 21 compétences.

Digital Promise, association internationale à but non lucratif, œuvrant pour le développement de la littératie de l'IA (*AI literacy*) a publié un référentiel en IA***. Elle part d'une définition de la littératie en IA :

« La littératie en intelligence artificielle englobe les connaissances et compétences permettant aux individus de comprendre, d'évaluer de manière critique et d'utiliser les systèmes et outils d'IA, afin de participer de façon sûre et efficace à un monde de plus en plus numérique. »

Les trois dimensions sont explicitées :

- **Comprendre** : Acquérir des connaissances de base sur ce que l'IA peut faire et comment elle fonctionne, afin de prendre des décisions éclairées concernant l'évaluation et l'utilisation des systèmes et outils d'IA.
- **Évaluer** : Mettre l'accent sur le jugement humain et la justice pour examiner de manière critique les avantages et/ou les coûts de l'IA pour les individus, la société et l'environnement.
- **Utiliser** : Interagir avec l'IA, créer et résoudre des problèmes grâce à elle, dans une progression d'usages adaptés à des contextes et objectifs spécifiques.



Ces trois dimensions inspirent largement le présent référentiel.

** [Vuorikari, R., Kluzer, S. and Punie, Y., DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes, EUR 31006 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-48883-5, doi:10.2760/490274, JRC128415.](#)

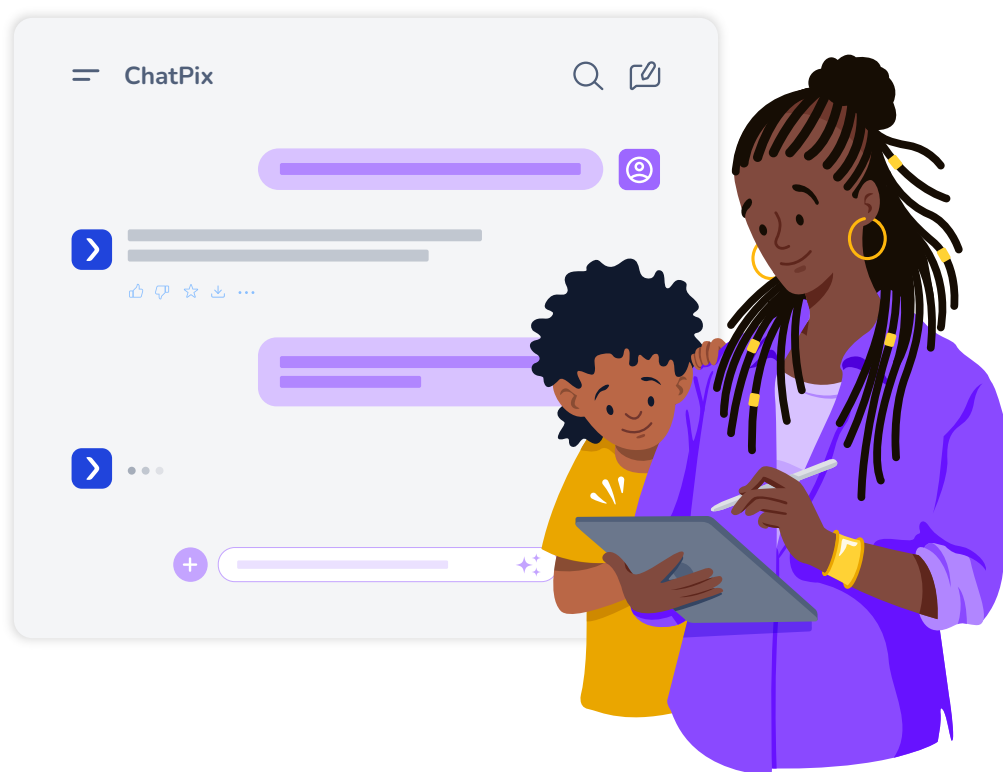
*** [Mills K., Ruiz P., Lee K.-W. Coenraad M., Fusco J., Roschelle J., Weisgrau J., AI Literacy: A Framework to Understand, Evaluate, and Use Emerging Technology, Digital Promise, June 2024](#)

Contours

L'intelligence artificielle réactualise des préoccupations déjà soulevées par l'utilisation massive des technologies numériques. Beaucoup des compétences numériques inscrites au **référentiel transverse de Pix** sont essentielles pour comprendre ce qu'est l'IA et les enjeux que soulèvent les utilisations des technologies qu'elle permet de développer.

Ainsi, savoir concevoir un algorithme et programmer permet de comprendre les algorithmes et modèles d'IA, savoir collecter, nettoyer, traiter et visualiser des données permet de comprendre le rôle des données dans les modèles d'IA et les enjeux concernant l'utilisation des données, connaître les matériels informatiques et leurs architectures permet de comprendre pourquoi les algorithmes d'IA ont un impact environnemental important, connaître les réglementations concernant l'utilisation des données, notamment les données personnelles, permet de comprendre pourquoi la collecte massive de données pour entraîner les IA renouvelle les enjeux éthiques et légaux, savoir évaluer l'information permet de juger de la qualité et de la pertinence des informations générée avec l'aide de système d'IA, etc.

Étant déjà décrites dans le référentiel transverse, ces compétences ne sont pas détaillées dans le présent document. Le présent référentiel mettra l'accent sur la façon dont l'IA complète et réinterroge ces compétences.



Sommaire

Avant-propos	2
------------------------	---

Domaines et compétences	7
-----------------------------------	---

Domaine 1 : Fondements de l'intelligence artificielle	7
---	---

- 1.1 Définir l'IA en tant que discipline, décrire son histoire et sa construction
- 1.2 Expliquer le procédé d'apprentissage automatique et ses approches
- 1.3 Citer les modèles de l'apprentissage automatique
- 1.4 Expliquer les méthodes d'entraînement des grands modèles de langage
- 1.5 Expliquer le fonctionnement des algorithmes de recommandation
- 1.6 Définir l'IA incarnée associée à la robotique

Domaine 2 : Usages et applications de l'IA	16
--	----

- 2.1 Identifier des familles de tâches réalisés par des logiciels d'IA
- 2.2 Utiliser un logiciel d'IA générative
- 2.3 Évaluer l'information à l'heure des IA génératives.
- 2.4 Utiliser les services à base d'algorithmes de recommandation
- 2.5 Utiliser des logiciels d'IA dans son organisation

Domaine 3 : Enjeux du développement de l'IA	22
---	----

- 3.1 Expliquer l'empreinte environnementale des supercalculateurs
- 3.2 Expliquer les enjeux de gouvernance des systèmes d'IA
- 3.3 Expliquer les enjeux éthiques et de transparence des modèles d'IA
- 3.4 Expliquer les conséquences sur l'emploi, la formation et la fracture numérique
- 3.5 Expliquer les enjeux culturels et sociétaux

Domaines et compétences

Ce référentiel décrit trois grands domaines découpés en compétences. Pour chaque compétence des exemples de savoirs et savoir-faire sont listés. Ces listes ne sont pas exhaustives, et les exemples ne relèvent pas toujours du même niveau de détail.

Domaine 1 : Fondements de l'intelligence artificielle

Comprendre les fondamentaux de l'intelligence artificielle (SIA) est essentiel pour en maîtriser les usages et en anticiper les impacts. Ce domaine porte sur la connaissance et la compréhension des sciences, des techniques, et des algorithmes du domaine de l'intelligence artificielle. Il couvre également la compréhension des principes de fonctionnement des systèmes d'IA les plus couramment rencontrés.

1.1

Définir l'IA en tant que discipline, décrire son histoire et sa construction

L'IA est un domaine interdisciplinaire qui s'est donné pour objectif de modéliser les mécanismes de l'**intelligence** humaine et animale pour mieux les comprendre, et de permettre à des machines d'effectuer des tâches complexes, éventuellement en **adaptant** leurs comportements au contexte ou à l'environnement.

Cet objectif a depuis toujours motivé la construction des machines par les humains. On date les débuts de la discipline au début des **années 50**, marquée par : la création du premier ordinateur à réseau de neurone, l'invention du test de Turing, l'adoption de l'expression « intelligence artificielle » à la conférence de Dartmouth. Dès ces débuts, les chercheurs ont misé sur la création d'ordinateurs dotés d'une capacité d'apprentissage et d'auto-adaptation permettant aux machines de faire plus qu'appliquer un comportement décrit pas à pas par un programme déterministe.

Deux grandes approches ont été explorées : des méthodes dites **symboliques**, fondées sur l'explicitation du raisonnement, et des méthodes **statistiques** basées sur un **apprentissage automatique** des machines à partir de données. On peut considérer que ces approches ont d'abord été développées en parallèle, mais qu'une démarche d'**hybridation** de ces méthodes a progressivement vu le jour.

La discipline a connu une alternance d'engouement médiatique, et de périodes de diminution de financement et d'intérêt (« les hivers de l'IA »). À partir des années 2000, la considérable augmentation des **capacités des ordinateurs** dans le traitement et le stockage de **grandes quantités de données** (« données massives » ou « bigdata »), ainsi que le déploiement de méthodes de **collecte de données à grande échelle** ont fortement contribué au progrès de l'apprentissage automatique. Les années 2020 ont été marquées par les avancées de l'**intelligence artificielle générative** et son déploiement dans des services accessibles à un grand nombre de personnes.

Ce déploiement grand public augmentant la visibilité des technologies conçues par cette discipline s'est accompagné d'un glissement de l'usage de l'expression « intelligence artificielle » pour désigner les technologies elles-mêmes voire même certains services ou logiciels spécifiques. Par exemple, des titres de presse désignent parfois par l'expression « une IA générative », tel ou tel agent conversationnel doté de capacités de génération de contenu.

Au fil du temps, différentes disciplines se sont construites et ont contribué au progrès de l'intelligence artificielle. L'**informatique** développe des infrastructures matérielles et logicielles nécessaires pour traiter et analyser de grandes quantités de données. Elle conçoit et améliore les algorithmes et les modèles d'apprentissage automatique. Les **sciences cognitives** apportent une compréhension approfondie des processus mentaux (perception, raisonnement, apprentissage, prise de décision) qui inspirent la conception de modèles et d'algorithmes d'IA. La **robotique** développe des technologies permettant d'associer les modèles d'IA à des architectures capables de percevoir et d'interagir avec le monde physique. Elle permet d'envisager le développement d'une **IA incarnée**. La **science des données** contribue à la conception des méthodes et outils nécessaires pour collecter, nettoyer, analyser et interpréter de grands ensembles de données. Elle développe des méthodes d'analyse statistique et d'exploration de données.

Cette histoire illustre les rôles conjoints de la démarche de tâtonnement scientifique, de l'interdisciplinarité, du soutien financier et de l'appropriation industrielle des technologies au développement de l'intelligence artificielle, en tant que discipline scientifique.

Exemples de savoirs et savoir-faire :

- Définir simplement l'IA, en tant que discipline scientifique (**novice**)
- Désigner l'augmentation de la capacité de calcul des ordinateurs et la collecte et le traitement de données massives comme facteurs de progrès de l'IA (**novice**)
- Expliquer le glissement du terme « intelligence artificielle » vers la désignation des technologies, ou des services et logiciels les incorporant (**indépendant**)
- Expliquer les contributions interdisciplinaires à l'IA (**avancé**)
- Différencier l'IA symbolique de l'apprentissage automatique, (**avancé**)
- Décrire les jalons historiques de l'IA, les événements scientifiques marquants et les « hivers » (**expert**)
- Décrire des exemples d'hybridation des approches symbolique et automatique (**expert**)

Mots clé :

#histoire, #discipline
scientifique, #intelligence,
#adaptation,
#IA symbolique,
#apprentissage
automatique, #capacité
ordinateur, #données
massives, #informatique,
#science cognitives

L'**apprentissage automatique** (machine learning en anglais) vise à donner aux machines la capacité d'apprendre un comportement à partir de **données**, via des **modèles mathématiques**.

Lors d'une **phase d'entraînement**, les informations pertinentes sont extraites d'un ensemble de données. Les paramètres de calcul du modèle sont ajustés afin qu'il atteigne les meilleures performances dans la réalisation de la tâche qui lui est attribuée. Une phase de test est nécessaire pour vérifier si le modèle fonctionne sur un jeu de données différent de celui de l'entraînement. Il s'agit d'évaluer la capacité de **généralisation** du modèle.

Différentes méthodes ont été développées selon le type de problème à résoudre et selon la nature des données dont on dispose : apprentissage supervisé, apprentissage non supervisé, apprentissage par renforcement.

L'**apprentissage supervisé** repose sur le principe de fournir au modèle des exemples de ce qu'il doit faire. Il s'agit de données d'entrée associées au résultat à produire ou **données étiquetées**. Le modèle cherche des tendances statistiques dans les données lui permettant de généraliser la décision à de nouveaux exemples. L'**apprentissage non supervisé** détecte des structures ou regroupements dans des données non étiquetées. Ces deux approches sont particulièrement adaptées pour apprendre à réaliser des tâches qui consistent à **prédire** une valeur.

L'**apprentissage par renforcement** consiste à décrire au modèle un environnement avec des propriétés et les actions possibles. Un **système de récompense** est mis en place pour donner un score à chaque action du modèle. Le modèle apprend par essais et erreurs pour maximiser son score. Cette approche est particulièrement adaptée pour apprendre à réaliser des tâches consistant à prendre une décision dans un **environnement** changeant.

Ces approches peuvent être **combinées**. Par exemple, un logiciel de jeu de Go, comme alphaGo, a d'abord appris les bases du jeu de Go, en s'entraînant sur des millions de positions de plateau de Go étiquetées par les mouvements joués par des experts humains (apprentissage supervisé) puis a amélioré ses compétences en jouant contre lui-même, par essai-erreur (apprentissage par renforcement). Autre exemple, un agent conversationnel avec des capacités génératives, comme ChatGTP, a appris à manier le langage en s'entraînant sur des milliards d'extrait de textes étiquetés par le mot suivant (apprentissage supervisé), puis a amélioré ses capacités de conversation en tenant compte, notamment, des retours des utilisateurs vue comme des récompenses (apprentissage par renforcement).

Il peut être nécessaire de ne pas laisser l'entraînement se réaliser de manière totalement automatique, en ajoutant des contraintes au modèle. En fait, cela revient à ajouter volontairement des **biais** à la sélection des résultats par le modèle. Selon le type d'apprentissage et de situation, cela peut être utile, par exemple, pour réduire la complexité du problème, améliorer la généralisation, introduire des connaissances a priori, guider l'apprentissage vers des solutions pertinentes, ou faire préférer au modèle certains choix afin de les aligner sur certains critères.

Exemples de savoirs et savoir-faire :

- Définir simplement la phase d'entraînement d'un modèle (**novice**)
- Expliquer le procédé de l'apprentissage supervisé (**indépendant**)
- Choisir les étiquettes de données dans un cas simple d'apprentissage supervisé (**indépendant**)
- Expliquer le procédé de l'apprentissage par renforcement (**avancé**)
- Identifier des critères de récompense adaptés dans un cas d'apprentissage par renforcement (**avancé**)
- Expliquer les différentes méthodes d'apprentissage automatique (**expert**)
- Choisir la technique d'IA adaptée en fonction de la nature des données exploitables (**expert**)
- Expliquer l'intérêt d'introduire des biais dans le modèle dans certaines situations (**expert**)

Mots clé :

#apprentissage automatique,
#entraînement, #jeu de données #généralisation,
#biais volontaire,
#supervisé, #non supervisé, #renforcement

1.3

Citer les modèles de l'apprentissage automatique

Les modèles d'intelligence artificielle sont des structures mathématiques ou informatiques conçues pour apprendre à partir de données et réaliser des tâches complexes.

Certains modèles reposent sur des **arbres de décision** qui segmentent les données selon des règles successives, sous forme d'un arbre. Chaque nœud représente une question et les branches correspondent aux réponses (oui/non). En descendant dans l'arbre, on arrive à une feuille qui donne la prédiction. Ces modèles sont intuitifs, faciles à visualiser, et peuvent gérer des relations non linéaires, mais ils peuvent devenir complexes et sensibles au surapprentissage si l'arbre est trop profond.

Certains modèles reposent sur des calculs de **régression** (linéaire, logistique, etc.), pour trouver une relation entre une valeur que l'on souhaite prédire et des valeurs que l'on observe. Ces modèles sont particulièrement adaptés lorsqu'on dispose de données bien structurées et que la relation entre les variables est relativement simple ou peut être approximée par une fonction mathématique explicite. Ils sont faciles à interpréter, ce qui permet de comprendre l'influence de chaque variable sur la prédiction.

Certains modèles reposent sur des calculs **probabilistes**, pour modéliser l'incertitude et les relations entre différentes variables. Par exemple, les modèles de Markov permettent de prédire l'évolution d'un système en fonction de son état actuel, en attribuant des probabilités aux différents événements possibles. Ces modèles sont particulièrement utiles quand les données sont incertaines ou partielles.

Certains modèles, inspirés du cerveau humain, sont appelés **réseaux de neurones**. Un réseau est composé de nombreuses unités simples appelées neurones artificiels, organisées en couches. Chaque neurone reçoit des informations, effectue un calcul, puis transmet un signal (le résultat de son calcul) aux neurones suivants. Lors du processus d'apprentissage, le réseau ajuste les poids qui déterminent à quel point les signaux reçus sont pris en compte par les neurones. Différents algorithmes d'optimisation ont été imaginés pour réaliser ces ajustements. Par exemple, un algorithme consiste à faire de petits ajustements pas à pas dans la bonne direction (descente de gradient). Ces modèles sont particulièrement puissants pour traiter des données non structurées et résoudre des problèmes complexes, mais leur fonctionnement interne est souvent difficile à interpréter, ce qui les rend parfois qualifiés de « boîtes noires ».

Les modèles à la base des logiciels d'IA générative, sont des réseaux de neurones. Par exemple, les transformateurs sont des réseaux de neurones spécialisés dans la compréhension et la génération de texte.

Exemples de savoirs et savoir-faire :

- Prédire une valeur à partir d'un droit de régression linéaire **(novice)**
- Appliquer un arbre de décision pour prédire une solution **(novice)**
- Décrire simplement le fonctionnement général des réseaux de neurones artificiels **(indépendant)**
- Définir l'objectif d'un calcul de régression **(indépendant)**
- Décrire à quelle fin un calcul de régression peut être utilisé dans l'entraînement d'un logiciel d'IA **(avancé)**
- Citer des applications dans lesquels un réseau de neurones est utilisé **(avancé)**
- Analyser les limites en termes d'interprétabilité des différents modèles **(expert)**
- Comparer différents types de modèles selon les tâches et les données **(expert)**

Mots clé :

#modèle apprentissage automatique, #arbre de décision, #réseau de neurone, #régression, #modèle probabiliste, #explicabilité

1.4

Expliquer les méthodes d'entraînement des grands modèles de langage

Aujourd'hui, de nombreux logiciels **d'intelligence artificielle générative** sont disponibles au grand public. Certains sont spécialisés sur un type de contenus (texte, images), d'autres sont spécialisés sur certains sujets. Mais les plus connus, comme LeChat ou chatGPT, sont assez généralistes et progressent rapidement vers de nouvelles capacités.

Leur capacité de génération de réponse à une demande formulées en langage naturelle tient à l'utilisation de **grands modèles de langages**. Ces modèles sont pré-entraînés sur de **grands volumes de textes** issus du Web, selon un principe d'apprentissage supervisé. Ce modèle utilise un algorithme appelé **transformateur**, qui transforme chaque mot en une représentation mathématique (vecteur), utilise un mécanisme d'attention pour décider quels mots sont importants, affine la compréhension du contexte en détectant des relations entre les mots (grammaire, sens, etc), et prédit, en sortie, les mots suivants les plus probables.

Après ce pré-entraînement, une phase **d'alignement** consiste à améliorer le résultat du modèle pour qu'il réponde mieux aux demandes formulées. Les ingénieurs fournissent au modèle des **exemples de bonnes conversations**, sélectionnées ou construites. C'est à cette étape que le modèle apprend à converser, à résumer, à écrire certains documents types (CV, lettre, etc.), etc. Enfin, une autre étape consiste à faire évaluer les résultats du modèle par des humains qui notent les réponses faites à leur demande. Cette étape applique un principe **d'apprentissage par renforcement** (reinforcement learning from human feedback en anglais).

Une fois entraîné, le modèle a la capacité de répondre à une demande formulée par un utilisateur (une **requête** ou prompt en anglais) en la découpant en unité de texte (*token*) pour calculer un contexte et choisir les mots suivants les plus probables pour construire une réponse. La réponse est bien formulée, cohérente avec le sujet traité et adaptée à la demande de l'utilisateur.

Le pré-entraînement permet au modèle d'apprendre des connaissances sur le monde, d'autant plus importantes que les textes sont nombreux, variés et sur tous les sujets. Toutefois, il est possible de fournir à un logiciel d'IA génératif déjà entraîné des sources supplémentaires sous forme de fichiers ou de bases de données à consulter, on parle de **génération augmentée par récupération** (*retrieval-augmented generation* en anglais). C'est un moyen d'apporter au modèle des connaissances sur un sujet spécifique, par exemple lié à une organisation, sans ré-entraînement.

Enfin, des capacités peuvent être « greffées » à la capacité de traitement de la langue, selon un procédé **d'appel à des services ou logiciels spécialisés**. Ainsi, avant de déclencher la création de la réponse, un programme analyse la demande formulée (le prompt ou requête) et repère éventuellement s'il est nécessaire de faire appel à d'autres capacités, comme la recherche d'information sur le Web, le calcul mathématique ou la programmation. Le programme de pré-analyse ajoute alors au prompt le mode d'emploi d'outils utilisables par requête en ligne.

L'ensemble de ce procédé de construction de la réponse ne recherche pas et ne garantit donc pas la **véracité** des informations fournies. Ces modèles n'ont pas (encore) la capacité de déterminer si ce qu'ils disent est vrai, et ne peuvent pas déterminer comment ils le savent. Ils n'ont pas de capacité de **métacognition** comme les humains. Il peut arriver que les textes utilisés à l'entraînement comportent des **erreurs**, que le modèle reproduit dans ces réponses. Il peut aussi arriver que le procédé de création de la réponse construise des **informations inventées**, que l'on appelle **hallucinations (ou confabulations)**. Toutefois, c'est cette capacité à créer des inventions qui leur permet d'être créatives. Il est donc crucial de décider à quelles fins on les utilise, et de choisir la qualité de la réponse attendue entre imaginaire ou vérité.

Exemples de savoirs et savoir-faire :

- Expliquer le principe de recherche du mot suivant le plus probable **(novice)**
- Repérer une hallucination sur un sujet simple **(novice)**
- Décrire les étapes d'entraînement d'un modèle de langage **(indépendant)**
- Décrire le rôle des humains dans le processus d'entraînement d'un modèle de langage **(indépendant)**
- Expliquer les sources d'erreur des IA génératives **(indépendant)**
- Décrire l'appel à des logiciels spécialisés **(avancé)**
- Décrire les principes de fonctionnement d'un transformateur **(expert)**

Mots clé :

#grand modèle
de langage,
#transformateur,
#pré-entraînement,
#alignement,
#renforcement
par des humains,
#hallucinations, #erreurs

1.5

Expliquer le fonctionnement des algorithmes de recommandation

Les algorithmes de recommandation sont conçus pour recommander à un utilisateur des contenus, des produits ou des services susceptibles de l'intéresser. Ils sont omniprésents dans les services numériques : plateformes de vidéo (YouTube, Netflix), de musique (Spotify), de commerce en ligne (Amazon), de réseaux sociaux (TikTok, Instagram) ou encore de recherche d'information (Google News, moteurs de recherche).

Leur objectif est de **filtrer l'information** pour prioriser ce qui pourrait avoir le plus de valeur ou d'intérêt pour un utilisateur donné, en se basant sur des données disponibles : son comportement passé, ses préférences exprimées, les choix d'utilisateurs similaires, ou les caractéristiques des contenus eux-mêmes. L'objectif est de **personnaliser** les recommandations et de maximiser la probabilité que l'utilisateur clique, regarde, achète ou interagisse avec les contenus recommandés.

Trois approches existent. Le **filtrage collaboratif** recommande des contenus à un utilisateur en s'appuyant sur les ressemblances de ses préférences avec celles d'autres utilisateurs ayant des goûts similaires. Ce principe nécessite de disposer des historiques d'interactions de nombreux utilisateurs et souffre du problème du démarrage à froid (quand on ne dispose pas de données d'historique suffisantes). Le **filtrage basé sur le contenu** propose des contenus similaires à ceux que l'utilisateur a déjà appréciés, en analysant leurs caractéristiques (thème, auteur, genre...). Il construit un profil utilisateur basé sur ses préférences passées.

Des **approches hybrides** combinent ces méthodes (collaboratif + contenu, par exemple) pour améliorer la qualité des recommandations et compenser les limites de chaque approche.

Les systèmes de recommandation utilisent des **modèles prédictifs** pour réaliser le filtrage. Ces modèles sont entraînés le plus souvent selon un principe **d'apprentissage supervisé**. Ils sont entraînés sur des **données historiques** où les caractéristiques des utilisateurs et des contenus sont associées (étiquetées par) à leurs interactions sur des contenus. Des transformateurs peuvent être utilisés pour annoter automatiquement les contenus.

Les algorithmes de recommandation influencent profondément l'accès à l'information. Ils peuvent favoriser **l'enfermement algorithmique** (ou « bulle de filtre »), en exposant surtout des contenus similaires à ce que l'utilisateur a déjà vu. Cela peut limiter la diversité de l'information, renforcer certains biais ou polariser les opinions.

Ils posent aussi des questions de transparence et de responsabilité : il est souvent difficile de comprendre pourquoi un contenu est recommandé, et les critères de recommandation ne sont pas toujours explicités.

Exemples de savoirs et savoir-faire :

- Distinguer recommandation personnalisée et non personnalisée **(novice)**
- Identifier les d'usages courant de recommandation personnalisée **(novice)**
- Identifier les types de données utilisés par les algorithmes de recommandation **(indépendant)**
- Expliquer le risque d'enfermement dans une bulle de filtre **(indépendant)**
- Expliquer le principe du filtrage collaboratif et du filtrage par contenu **(avancé)**
- Expliquer le principe du démarrage à froid et ses conséquences **(expert)**
- Décrire le recours à l'apprentissage automatique dans les systèmes de recommandation **(expert)**
- Analyser les effets d'un algorithme de recommandation sur les comportements d'un utilisateur ou d'un groupe **(expert)**

Mots clé :

#recommandation,
#filtrage,
#personnalisation,
#filtrage collaboratif,
#historiques, #profil,
#bulle de filtre

1.6

Définir l'IA incarnée associée à la robotique

L'intelligence artificielle n'est pas toujours confinée à des logiciels ou à des services numériques : elle peut aussi être incarnée dans des **objets physiques** capables de percevoir, agir et interagir dans le monde réel. On parle alors d'**IA incarnée**, souvent à travers la **robotique**.

Un **robot** est un système composé de trois grandes fonctions : une **capacité à percevoir** son environnement (capteurs, caméras, micros, etc.), une **capacité à prendre des décisions** (souvent grâce à l'IA), une **capacité à agir** (moteurs, bras articulés, déplacements, etc.).

Quand un robot utilise des modèles d'intelligence artificielle pour comprendre son environnement, adapter ses actions ou apprendre de nouvelles tâches, on parle de **robot intelligent**. Par exemple, un robot peut utiliser la vision par ordinateur pour reconnaître un objet, la planification automatique pour décider comment le saisir, et l'apprentissage automatique pour améliorer ses gestes au fil du temps.

L'IA incarnée permet donc de faire le lien entre **perception, décision et action** dans le monde réel. Cela représente un défi particulier : le monde physique est **incertain, changeant, souvent bruyant**. Les robots doivent être capables de s'adapter rapidement, de réagir à l'imprévu, et de fonctionner en interaction avec des humains.

Exemples de savoirs et savoir-faire :

- Définir ce qu'est un robot et en identifier les grandes fonctions (**novice**)
- Donner des exemples de robots utilisant l'IA (**novice**)
- Expliquer ce qu'on appelle une IA incarnée (**indépendant**)
- Identifier les enjeux liés à l'incertitude du monde réel dans l'IA incarnée (**indépendant**)
- Expliquer comment un robot combine perception, décision et action (**avancé**)

Mots clé :

#IA incarnée, #robot, #perception, #action physique, #interaction



Domaine 2 : Usages et applications de l'IA

2.1

Identifier des familles de tâches réalisés par des logiciels d'IA

Les tâches que les machines (ordinateurs ou robots) peuvent réaliser grâce à l'intelligence artificielle sont variées. Les humains réalisent certaines de ces tâches sans y penser comme les tâches de perception (voir, entendre), ou des tâches d'interaction (parler, bouger, toucher). Pour d'autres tâches, les humains ont besoin de plus de concentration, comme des tâches de classement, de prédiction, de prise de décision, de création, etc. Pour les machines, toutes ces tâches sont réalisées grâce à des modèles mathématiques ou informatiques nécessitant une étape d'entraînement.

La **reconnaissance d'images** ou de vidéos consiste à analyser des images ou des vidéos pour en extraire des informations. Il s'agit par exemple de reconnaître certains objets, ou des visages, ou des gestes, ou même d'analyser des scènes complexes. La **reconnaissance de sons** consiste à traiter des signaux audio pour détecter des sons, ou reconnaître la parole, ou transcrire un discours en texte. Ces tâches reposent sur des traitements spécifiques du signal visuel ou sonore, produisant des données utilisées en entrée de modèles d'apprentissage automatique. La reconnaissance d'image est utilisée par les applications de photographies qui font de la retouche d'image automatique, par les logiciels embarqués sur une voitures autonome pour reconnaître les obstacles sur son chemin (piétons, vélos, etc), par des logiciels médicaux d'aide à la décision pour détecter des tumeurs sur des images médicales, etc.

La **prédiction** consiste à estimer une valeur ou un événement futur à partir de données passées. Il peut s'agir de prévoir des valeurs numériques, des comportements ou des événements. La prédiction a beaucoup d'utilisations pour anticiper des événements et prendre une décision. Par exemple prédire les pannes de machines, prédire le trafic routier, prédire des comportements frauduleux, prédire un risque de maladie, etc.

La **recommandation** personnalisée de contenus à un utilisateur est également une tâche permise par l'IA. Les systèmes de recommandation visent à filtrer, ordonner ou proposer automatiquement des éléments jugés pertinents. Pour ce faire, ils prennent en compte les préférences ou le comportement de l'utilisateur et, soit les préférences ou le comportement d'autres utilisateurs similaires, soit les caractéristiques des contenus similaires à ceux déjà choisis par l'utilisateur. L'IA est nécessaire pour entraîner les systèmes sur de grandes quantités de données. La recommandation personnalisée est utilisée dans les services de vidéo ou de chansons à la demande, sur les réseaux sociaux, sur certains sites de presse, etc.

La **génération de contenus** consiste à créer de nouveaux contenus (textes, images, sons, musiques, ou vidéos) à partir d'une consigne. Les logiciels de génération de contenu sont capables de converser avec un humain sur des sujets variés, trouver de nouvelles idées, créer des contenus inventifs. Ils peuvent être augmentés par la capacité de se connecter à des logiciels existants par exemple pour faire une recherche d'information sur le web, faire des calculs, programmer, mais aussi déclencher des actions sur des systèmes d'informations (de banque, d'entreprise, etc). Ils deviennent de plus en plus polyvalents. La génération de contenus est utilisée dans les assistants conversationnels grand public accessibles en ligne, comme ChatGPT, Le chat, ou intégrés dans des suites bureautiques comme Gemini ou Copilot.

Exemples de savoirs et savoir-faire :

- Identifier les utilisations des techniques d'intelligence artificielle dans des outils et services utilisés au quotidien **(novice)**
- Expliquer la finalité d'un système de reconnaissance visuelle ou sonore **(novice)**
- Identifier différentes familles d'applications de l'IA **(indépendant)**
- Identifier les applications présentes dans des outils utilisés quotidiennement. **(indépendant)**
- Donner des exemples concrets d'utilisation de l'IA pour la prédiction dans différents domaines **(avancé)**
- Expliquer la polyvalence des assistants à base d'IA générative **(avancé)**

Mots clé :

#familles d'applications,
#IA au quotidien,
#reconnaissance d'image,
reconnaissance de son,
#tâche de prédiction,
#recommandation,
#génération de contenu

2.2 Utiliser un logiciel d'IA générative

Les logiciels d'IA générative sont utiles pour rédiger des textes et différents documents types, créer des images ou des vidéos, résumer ou traduire ou analyser des textes, trouver des informations, consulter une base de données ou de documents, écrire des programmes, faire des calculs, développer des raisonnements, etc et des combinaisons de tout cela.

Le mode d'interaction avec le logiciel consiste à écrire une **requête en langage naturel**. Plus la requête est précise et contextualisée, plus la réponse répondra au besoin.

Des **pratiques d'écriture** permettent de formuler correctement la requête : utiliser des verbes d'action (rédige, résume, dessine, programme, calcule, classe, cherche, etc), préciser l'objectif, donner tous les éléments de contexte, donner des éléments concernant la forme (le ton ou le format de présentation dans le cas d'un texte, le style pictural pour une image, etc). Il est possible de fournir avec la requête des fichiers qui apportent des informations ou des données supplémentaires au logiciel.

Certains verbes associés à certains éléments de contexte peuvent **déclencher l'appel** à des logiciels spécialisés dans le calcul, la recherche d'information, la programmation, etc. (Par exemple : « réalise le calcul de probabilité de », « trace la courbe de régression de », « crée une visualisation montrant la répartition de », « donne une source pour chaque », etc.)

Certaines formulations permettent de déclencher un mode de fonctionnement spécifique. Par exemple, préciser dans le prompt « procède par étapes successives », peut déclencher une construction de la réponse étape par étape, avec des justifications intermédiaires (**chaîne de pensée**). Dans certains cas, cela peut améliorer la fiabilité de la solution (comme des exercices de logique ou de mathématique).

Il est souvent utile de procéder par **itérations successives** avec le logiciel, en lui faisant modifier, compléter, développer, ou autre sa réponse précédente. Il est toujours nécessaire de vérifier le résultat obtenu. Il peut contenir des faits imaginaires présentés comme réels (appelées **hallucinations**). Il peut aussi contenir des **erreurs** de raisonnement.

En général, les contenus des discussions peuvent être utilisés pour entraîner le logiciel d'IA générative. C'est-à-dire que toutes les données fournies peuvent être apprises par le logiciel, et éventuellement utilisées dans une conversation ultérieure avec d'autres utilisateurs. Il est prudent de ne pas fournir de **données personnelles, sensibles, confidentielles, ou sous droit d'auteur**.

Exemples de savoirs et savoir-faire :

- Ecrire des requêtes appliquant des bonnes pratiques de clarté et précision (**novice, indépendant**)
- Ajuster et affiner la requête en fonction des résultats obtenus (**indépendant, avancé**)
- Identifier les types de tâches qu'on peut demander à une IA générative (résumer, comparer, développer, réécrire dans un autre style, traduire, etc.) (**indépendant**)
- Évaluer la réponse d'une IA générative, pour détecter d'éventuelles erreurs ou hallucinations non souhaitées (**novice, indépendant**)
- Interroger une IA générative avec différentes modalités (**avancé**)
- Expérimenter différents types de requêtes pour explorer les capacités et les limites des modèles d'IA générative (**avancé**)
- Vérifier la réponse d'une IA générative en croisant avec d'autres sources d'information (**avancé**)

Mots clé :

#IA générative, #requête, #règles d'écriture, #chaîne de pensée, #vérification, #itération

2.3

Évaluer l'information à l'heure des IA génératives.

Les IA génératives sont largement utilisées pour produire des contenus truqués (**hypertrucages** ou **deepfake**) qui sont publiés sur le Web et les réseaux sociaux.

La production d'hypertrucage demande peu de compétences techniques et peu d'éléments d'entrée. Une photo d'une personne et quelques phrases prononcées par elle suffisent pour créer une vidéo de cette personne disant un texte qu'elle n'a jamais prononcé. Il est souvent très **difficile de détecter** à l'œil nu qu'un contenu a été produit par ou avec une technique d'IA. Les techniques avancées de détection de manipulation algorithmique, notamment dans les images, ne parviennent plus à détecter un hypertrucage.

Des robots logiciels (**bots**) sont utilisés pour produire **automatiquement** des textes, images ou vidéos à l'aide d'outils d'IA générative et de les publier de **façon massive** sur des plateformes sociales, parfois sans intervention humaine directe. Ils peuvent publier un contenu sur une chaîne, poster des commentaires sous des publications, poster des messages dans des forums, plussoyer des contenus, etc. Ainsi, ils sont capables de simuler des opinions, diffuser des informations erronées, ou faire apparaître des sujets comme populaires alors qu'ils ne le sont pas.

La diffusion de ces contenus est amplifiée par les **algorithmes de recommandation**, qui favorisent les contenus qui font le plus réagir (aimer, consulter, republier, etc) les utilisateurs. Or les contenus les plus **attirants** pour les humains sont souvent des contenus sensationnels, erronés, négatifs quel que soit leur degré de véracité. Ainsi, des contenus trompeurs, générés automatiquement ou pas, peuvent se retrouver rapidement exposés à un grand nombre de personnes, renforçant les risques de désinformation.

À titre individuel, les méthodes classiques de **vérification de l'information** (*fact checking*) restent la meilleure parade : déterminer qui est l'auteur, quelles sont ses intentions, s'il a lui-même vérifié les informations qu'il communique, vérifier la présence de ses informations dans d'autres publications. Il est également important de **signaler** l'hyper trucage sur le site ou le réseau où il est diffusé, afin d'encourager les plateformes à vérifier ces informations et éventuellement les dépublier.

Des pratiques de **mention** par l'auteur qu'un contenu a été produit avec l'aide d'un logiciel d'IA et des pratiques de **marquage** invisible des contenus par l'IA qui les produit se développent mais aucune loi ne garantit qu'elles soient utilisées.

Exemples de savoirs et savoir-faire :

- Identifier l'intention de l'auteur d'un contenu truqué trouvé en ligne (**novice**)
- Interpréter une mention d'utilisation de l'IA sur un contenu en ligne (**novice**)
- Trouver l'auteur d'un contenu publié en ligne (**indépendant**)
- Distinguer la méthode d'une IA générative avec celle d'un moteur de recherche (**indépendant**)
- Expliquer comment les algorithmes de recommandation peuvent favoriser des contenus truqués postés par des bots (**avancé**)
- Appliquer des principes de vérification de fait pour vérifier une information (**avancé**)

Mots clé :

#hypertrucage, #bots, #IA générative, #algorithmes de recommandation, #vérification de faits, #mention, #marquage

De nombreux services numériques — plateformes de vidéos, réseaux sociaux, sites d'achats en ligne — utilisent des **systèmes de recommandation** pour sélectionner les contenus qu'ils proposent. Cette **personnalisation** repose sur les traces laissées en ligne : vidéos regardées, clics, recherches effectuées ou produits consultés. L'objectif est de rendre l'expérience plus fluide et adaptée aux préférences de chaque utilisateur.

Cette capacité à ajuster les suggestions aux goûts de chacun présente des **avantages**. Elle offre un filtrage personnalisé dans une offre vaste de contenus. Elle permet de découvrir plus facilement de nouveaux contenus, de gagner du temps et d'accéder à des propositions jugées pertinentes.

Cependant, ces systèmes peuvent aussi avoir des effets moins visibles, comme le phénomène d'**enfermement algorithmique**. Se voyant proposer en priorité ce qui correspond déjà à ses habitudes ou à ses centres d'intérêt, un individu pourrait ne plus être exposé à des idées différentes des siennes, à des contenus différents de ceux qu'il consulte déjà. On désigne par **chambre d'écho** la situation où la personne est exposée presque uniquement à des opinions similaires aux siennes, renforçant encore ces idées sans confrontation à des points de vue différents. Ce phénomène serait d'autant plus fort que l'individu a des goûts tranchés ou des opinions extrêmes.

Certaines plateformes offrent des **paramètres** permettant de mieux comprendre ou de moduler cette personnalisation : consultation de l'historique, modification des préférences enregistrées, ou encore désactivation de certaines options. Ces outils permettent de garder un certain contrôle sur les contenus proposés, mais ils ne sont pas toujours faciles d'accès, et pas toujours proposés.

Adopter une posture de curiosité reste essentiel. Explorer d'autres types de contenus que ceux proposés automatiquement, diversifier les sources ou élargir ses recherches permet de sortir des logiques de répétition. Les algorithmes facilitent l'accès à ce qui plaît, mais ne remplacent pas la liberté de chercher par soi-même.

Exemples de savoirs et savoir-faire :

- Identifier les services et plateformes utilisant des algorithmes de recommandation (**novice**)
- Identifier les avantages et les limites d'un service utilisant la recommandation personnalisée (**indépendant**)
- Repérer les options de gestion de l'historique ou de la personnalisation dans un service numérique (**indépendant**)
- Expliquer en quoi la personnalisation peut limiter la diversité des contenus consultés (**indépendant**)
- Modifier les paramètres d'un compte pour gérer les recommandations affichées (**avancé**)
- Mettre en œuvre des pratiques de diversification de contenus pour éviter l'enfermement algorithmique (**avancé**)

Mots clé :

#algorithme de recommandation, #filtre personnalisé, #enfermement algorithmique, #bulle de filtre, #chambre d'écho, #paramétrage, #curiosité

L'intelligence artificielle peut être un levier efficace pour améliorer certaines activités au sein d'une organisation, qu'elle soit professionnelle, associative ou institutionnelle. L'IA est déjà utilisée dans des outils variés, souvent **intégrés dans des logiciels** de bureautique, de gestion de projet, de communication, de conception ou de création. Elle peut automatiser certaines tâches, aider à la prise de décision, suggérer des contenus, ou encore assister dans la production de documents ou d'analyses.

Utiliser l'IA dans un cadre organisationnel demande **d'identifier les besoins** : gagner du temps sur des tâches répétitives, améliorer la qualité d'un service, générer des contenus, analyser des données, soutenir la relation client ou renforcer la veille d'information. Ensuite, il faut **repérer les outils** qui embarquent une forme d'intelligence artificielle et dont les fonctionnalités correspondent à l'objectif visé. Il peut s'agir, par exemple, d'outils de rédaction assistée, de traduction automatique, d'analyse d'opinion, de génération d'images, ou de classification automatique de messages.

Certains outils exploitent des modèles d'IA **générative** capables de produire des textes, des synthèses ou des réponses sur la base d'informations fournies en entrée. Pour enrichir leur pertinence dans un contexte professionnel, on peut recourir à des systèmes de **génération augmentée par récupération** (RAG, *Retrieval-Augmented Generation* en anglais). Cette approche combine un moteur de recherche interne (souvent sur la documentation propre à l'organisation) avec un modèle de génération de texte. L'IA va d'abord extraire les informations les plus pertinentes d'une base de connaissances ou d'un corpus défini, puis générer une réponse en s'appuyant directement sur ces sources. Cela permet de réduire les hallucinations et d'obtenir des réponses alignées avec la documentation ou les procédures internes de l'organisation.

Le choix des outils implique aussi de vérifier certains éléments : le respect de la confidentialité des données, les conditions d'utilisation, la simplicité d'intégration dans l'environnement numérique de l'organisation, les limites techniques ou éthiques du service. Il est important de lire les **mentions légales** et les paramètres liés à la **collecte des données**, surtout lorsque les contenus traités sont sensibles ou internes. Pour encadrer l'usage de ces outils, certaines organisations mettent en place une **charte interne** définissant les règles d'utilisation de l'IA. Cette charte peut préciser les types de tâches autorisées, les précautions à prendre en matière de données personnelles ou confidentielles, et les responsabilités des utilisateurs. Elle vise à instaurer un usage réfléchi et éthique de l'IA, adapté aux objectifs et aux valeurs de la structure.

Exemples de savoirs et savoir-faire :

- Identifier des outils intégrant de l'IA **(novice)**
- Vérifier qu'un cas d'usage respecte une charte interne d'usage de l'IA **(novice)**
- Lister les enjeux liés à la confidentialité des données et aux conditions d'utilisation des outils d'IA **(indépendant)**
- Vérifier les paramètres et mentions légales liés à un outil d'IA avant son adoption **(indépendant)**
- Expliquer le principe de la génération augmentée par récupération (RAG) et son intérêt **(avancé)**
- Expliquer l'apport d'une charte interne d'usage de l'IA **(avancé)**
- Utiliser un système de génération augmentée par récupération pour obtenir des réponses fiables dans un contexte professionnel **(avancé)**

Mots clé :

#logiciels, #besoin,
#charte d'usage,
#génération augmentée,
#confidentialité, #mention
légale, #données
sensibles, #données
personnelles

Domaine 3 : Enjeux du développement de l'IA

3.1

Expliquer l'empreinte environnementale des supercalculateurs

La conception, le développement et l'utilisation massive de l'intelligence artificielle participent à l'augmentation de l'**empreinte environnementale** du secteur numérique. Pour évaluer cette empreinte, il est nécessaire de mobiliser les concepts et méthodes utilisés pour mesurer l'impact environnemental du numérique dans son ensemble (notamment la notion d'empreinte carbone), adaptés aux spécificités des technologies d'IA.

L'**entraînement** des modèles d'IA générative, en particulier les grands modèles utilisés pour le traitement du langage naturel, la reconnaissance d'images ou d'autres tâches complexes, demande d'**importantes capacités de calcul**. Ces calculs sont réalisés sur des infrastructures informatiques massives, telles que des fermes de serveurs ou des **supercalculateurs**, qui consomment beaucoup d'électricité, non seulement pour le calcul lui-même, mais aussi pour le **refroidissement** des équipements. Cette consommation énergétique, souvent fournie par des sources non renouvelables, contribue significativement aux émissions de gaz à effet de serre.

En amont, la **fabrication du matériel** nécessaire au fonctionnement de ces infrastructures — cartes graphiques, processeurs spécialisés, serveurs, et autres composants électroniques — fait appel à l'extraction de **ressources naturelles rares**, telles que le lithium, le cobalt, ou les terres rares. L'exploitation de ces matériaux a des conséquences néfastes sur l'environnement : pollution, consommation excessive d'eau, déforestation. Elles ont également des conséquences sociales dans les zones d'extraction.

Par ailleurs, l'**utilisation intensive** de certains services basés sur l'IA, notamment l'IA générative accessible au grand public, entraîne une augmentation rapide des demandes de calcul, amplifiant ainsi la consommation énergétique globale. L'engouement pour ces technologies conduit à une multiplication des usages, parfois sans évaluation préalable des bénéfices réels par rapport à leur coût environnemental.

Mesurer précisément la consommation énergétique et l'impact environnemental des logiciels d'IA générative reste une tâche complexe. Les infrastructures utilisées sont souvent réparties à travers le monde, rendant difficile le suivi exact de leur consommation électrique. De plus, les organisations ne communiquent pas toujours de manière **transparente** sur la quantité d'énergie utilisée, ni sur l'empreinte carbone générée par leurs activités d'IA. Les indicateurs de consommation (kilowattheures (kWh), émissions de CO2 équivalent, nombre d'opérations de calcul) sont souvent **estimées** et manquent de standardisation, ce qui complique les **comparaisons** entre différents modèles, outils ou organisations.

Face à ces enjeux, plusieurs pistes sont explorées pour réduire ses impacts environnementaux. L'**optimisation des algorithmes** pour qu'ils nécessitent moins de calculs — souvent appelée « IA frugale » ou « IA économe » — vise à rendre les modèles plus efficaces sans sacrifier leur performance. L'adoption de **sources d'énergie renouvelable** pour alimenter les centres de données est également encouragée afin de diminuer l'empreinte carbone. Par ailleurs, la **mutualisation des modèles d'IA**, via le partage et la réutilisation, permet d'éviter la duplication coûteuse en ressources des phases d'entraînement.

Enfin, il est possible d'adopter une **démarche réfléchie** dans le déploiement et l'utilisation de l'IA, en ciblant prioritairement les usages qui apportent une réelle valeur ajoutée sociétale, économique ou environnementale, plutôt que de promouvoir une expansion systématique et généralisée.

Exemples de savoirs et savoir-faire :

- Expliquer que les supercalculateurs qui hébergent des IA génératives grand public consomment beaucoup d'énergie pour traiter de grands nombres de requêtes **(novice)**
- Expliquer que l'entraînement d'une IA générative consomme beaucoup d'énergie **(indépendant)**
- Décrire les principales pistes pour réduire l'impact environnemental de l'IA **(indépendant)**
- Comparer différents logiciels d'IA générative selon leur efficacité énergétique **(indépendant)**
- Expliquer que la fabrication des composants des supercalculateurs utilise des ressources naturelles rares. **(indépendant)**
- Identifier les sources d'émission de gaz à effet de serre directe et indirecte dans le cycle de vie des modèles d'IA **(avancé)**
- Comparer la consommation de ses usages d'IA générative à d'autres pratiques numériques et non numériques **(avancé)**
- Expliquer le concept d'IA frugale **(avancé)**

Mots clé :

#supercalculateur
#centres de calcul,
#infrastructure, #IA
générative, #IA frugale,
#consommation
énergétique,
#ressources rares,
#empreinte carbone,
#énergie renouvelable,
#mutualisation

3.2

Expliquer les enjeux de gouvernance des systèmes d'IA

Le développement et le déploiement de l'intelligence artificielle ne posent pas seulement des questions techniques, mais soulèvent des enjeux profondément politiques. En effet, ils engagent des choix de société, de valeurs, de priorités collectives.

L'un des enjeux politiques majeurs de l'IA concerne la **souveraineté numérique**. La plupart des grandes plateformes d'IA sont aujourd'hui contrôlées par un nombre restreint d'entreprises situées dans quelques pays. Cette concentration de puissance technique et économique interroge la capacité des États, des organisations publiques ou de la société civile à garder le contrôle sur des outils aussi influents. Elle pose la question de l'indépendance technologique et de la capacité à imposer des normes adaptées à chaque contexte.

La **gouvernance de l'IA** est également un enjeu central. Qui décide de ce que l'IA doit faire ou non ? Qui définit les règles, les limites, les obligations en matière de transparence ou de respect des droits ? Face à ces questions, plusieurs approches coexistent : certaines insistent sur l'auto-régulation par les entreprises, d'autres sur le besoin d'une régulation publique forte, au niveau national, européen ou international. L'élaboration de lois spécifiques, comme le **règlement européen** sur l'intelligence artificielle (IA Act), témoigne d'une volonté croissante d'encadrer ces technologies.

Les **valeurs encodées** dans les systèmes d'IA sont elles aussi politiques. Lorsqu'une IA est conçue pour filtrer des informations, noter des profils, ou générer des contenus, elle repose sur des choix implicites ou explicites : ce qui est acceptable ou non, ce qui est pertinent, ce qui est vrai, juste, ou souhaitable. Ces choix traduisent une vision du monde, et peuvent renforcer certaines normes ou exclusions. Le développement de l'IA dite « **constitutionnelle** » (ou constitutional AI) reflète cette prise de conscience : il s'agit d'intégrer dès la conception des principes éthiques ou juridiques clairs, mais cela suppose de décider collectivement quelles valeurs l'IA doit respecter — et pour qui elle est conçue.

Enfin, l'IA s'inscrit dans un contexte **géopolitique** de compétition entre grandes puissances, qui investissent massivement pour développer leurs propres modèles, infrastructures et normes. Cela crée un déséquilibre dans l'accès aux technologies, dans la capacité à influencer les standards internationaux, et dans la répartition des bénéfices de l'innovation. L'intelligence artificielle devient ainsi un levier stratégique, non seulement pour des raisons économiques, mais aussi pour la sécurité, la défense ou l'influence culturelle.

Exemples de savoirs et savoir-faire :

- Savoir que l'IA fait l'objet de régulations à différentes échelles (entreprises, États, organisations internationales) **(novice)**
- Identifier des acteurs internationaux influents dans le domaine de l'IA **(indépendant)**
- Illustrer la notion de valeurs encodées dans une IA générative **(indépendant)**
- Expliquer ce que signifie « gouverner » une technologie comme l'IA **(indépendant)**
- Évaluer un cas d'utilisation de l'IA selon la grille de l'IA Act **(indépendant)**
- Lister les grandes lignes des cadres de gouvernance existants (ex : AI Act européen) **(avancé)**

3.3

Expliquer les enjeux éthiques et de transparence des modèles d'IA

Les enjeux éthiques concernent à la fois la conception, l'usage et la régulation des systèmes algorithmiques. L'éthique des algorithmes repose sur plusieurs principes fondamentaux : la **transparence** (comprendre comment fonctionne un système et sur quoi il se base), l'**explicabilité** (être capable d'expliquer ses décisions ou résultats), la **non-discrimination** (éviter que des biais présents dans les données ou les modèles ne (re-)produisent des inégalités), et la **justice** (veiller à ce que l'IA soit utilisée de manière équitable et responsable). Ces principes devraient guider le développement des systèmes et encadrer leur déploiement.

Dans certains cas, des **dilemmes éthiques** peuvent émerger, notamment concernant l'arbitrage entre performance et transparence. De même, certaines applications (comme les systèmes de notation sociale, les usages militaires ou la reconnaissance faciale à large échelle) posent des questions sur les libertés individuelles, l'autonomie humaine et le respect des droits.

La réglementation européenne de l'intelligence artificielle (**IA Act**) classe les systèmes d'IA selon le niveau de risque de leur utilisation (minime, limité, élevé, interdit) et impose des obligations spécifiques, en particulier pour les usages sensibles comme les soins médicaux, la justice ou l'éducation. En parallèle, certaines lois nationales évoluent pour intégrer les spécificités de l'IA dans les questions de **responsabilité juridique**, notamment en cas d'erreur ou de dommage causé par une décision algorithmique.

Un autre levier de régulation est l'**audit des algorithmes**. Il s'agit d'évaluer, de manière indépendante et méthodique, le fonctionnement d'un système d'IA pour détecter d'éventuels biais, vérifier sa conformité aux règles ou principes établis, et assurer une transparence vis-à-vis des utilisateurs et des parties prenantes.

Le libre accès aux données d'entraînement permet à des chercheurs, des entreprises, des administrations ou des citoyens de créer ou d'améliorer des systèmes d'IA sans dépendre uniquement de bases de données propriétaires. Les modèles à **poids ouverts** (open weight en anglais), désigne un modèle d'intelligence artificielle dont les paramètres appris pendant l'entraînement sont rendus publics. Des plateformes sont spécialisées dans la diffusion de ces modèles (ex : Hugging Face, GitHub, Model Zoo) En publiant ces poids, les créateurs permettent à d'autres de **reproduire, modifier, comprendre** ou **réentraîner** le modèle, sans repartir de zéro. Par exemple, des modèles comme LLaMA (Meta), Mistral, ou Falcon ont des poids disponibles sous certaines **licences**. Cela favorise l'innovation, la **transparence**, la recherche indépendante, et l'autonomie technologique, notamment dans les pays ou les organisations qui ne peuvent pas développer de grands modèles seuls.

Exemples de savoirs et savoir-faire :

- Illustrer un cas où l'erreur d'un système d'IA poserait la question de la responsabilité juridique (**novice**)
- Évaluer des situations d'utilisation de système d'IA au regard des cadres législatifs, comme le RGPD ou le l'AI Act (**indépendant**)
- Trouver un modèle à poids ouverts sur une plateformes (**avancé**)
- Expliquer la différence entre code source ouvert, données ouvertes et poids ouverts (**avancé**)
- Différencier les types de modèles selon leur degré d'explicabilité (**avancé**)

Le développement de l'IA soulève des enjeux sociaux et économiques, touchant à la fois l'emploi, les conditions de travail, la répartition des opportunités et les nouvelles formes d'inégalités.

L'automatisation permise par l'IA modifie la nature de nombreux métiers. Certaines tâches sont entièrement prises en charge par des systèmes automatisés, ce qui peut entraîner la disparition ou la transformation d'emplois. Dans le même temps, de nouvelles fonctions émergent, liées à la conception, à la supervision, à l'analyse ou à l'usage de ces technologies. Cela suppose une évolution des compétences dans presque tous les secteurs, avec un besoin accru de formation tout au long de la vie. La question n'est donc pas uniquement celle de la suppression d'emplois, mais celle du **déplacement du travail**, des compétences attendues et de la capacité d'adaptation des systèmes éducatifs et de formation.

Derrière les performances des intelligences artificielles se cachent souvent des milliers d'heures de **travail humain**, peu visibles. L'entraînement des modèles nécessite l'étiquetage, la validation ou la correction de grandes quantités de données – images, textes, vidéos – effectués par des humains. De même, la modération de contenu pour les plateformes repose largement sur des humains chargés de visionner et filtrer des contenus parfois choquants. On parle de **travailleurs du clic**, qui réalisent des tâches répétitives, s'exerçant souvent dans des conditions précaires, avec de faibles rémunérations, peu de reconnaissance ou de protection.

Les disparités géographiques, économiques et éducatives influencent fortement la capacité à utiliser ou à bénéficier de ces outils. Certaines populations ou régions n'ont pas accès aux infrastructures numériques nécessaires (connexion, matériel), d'autres manquent de formation ou d'accompagnement pour tirer parti de ces technologies. Cette **fracture numérique** peut accentuer des inégalités existantes et créer un écart entre ceux qui maîtrisent les outils d'IA et ceux qui en restent à l'écart.

Exemples de savoirs et savoir-faire :

- Citer des exemples de métiers qui disparaissent du fait de l'IA (**novice**)
- Citer des exemples de métiers qui apparaissent du fait de l'IA (**indépendant**)
- Repérer les tâches humaines nécessaires à l'entraînement ou à l'usage d'un modèle d'IA (**indépendant**)

Les technologies de l'IA ne sont pas neutres : elles reflètent, reproduisent et parfois renforcent les normes, les représentations et les inégalités de nos sociétés.

Les systèmes d'IA apprennent à partir de données, souvent collectées sur le web ou dans des bases existantes. Or, ces données peuvent contenir des **discriminations et stéréotypes** déjà présents dans la société : sexisme, racisme, stéréotypes de classe ou d'âge, etc. Par exemple, un modèle de langage peut associer certaines professions à un genre particulier. Ces biais sont souvent involontaires, mais ont des conséquences concrètes : ils peuvent influencer des recommandations, des décisions ou des évaluations automatiques, et contribuer à la reproduction des inégalités.

L'**enfermement algorithmique** constitue un autre enjeu. Les algorithmes de recommandation, en particulier sur les plateformes de contenus, tendent à proposer des informations similaires à celles que l'utilisateur a déjà vues ou appréciées. Cela favorise la formation de **chambres d'écho**, où chacun est exposé en priorité à des opinions proches des siennes, au détriment de la diversité des points de vue. Ce phénomène peut renforcer les **biais de confirmation** – la tendance à chercher et croire des informations qui confirment ses opinions – et contribuer à la **polarisation** des débats publics, avec un affaiblissement du dialogue démocratique.

Enfin, les technologies d'IA jouent un rôle dans la **manipulation de l'information**, à travers la génération automatique de contenus, l'amplification algorithmique de messages, la création de faux contenus (hyper trucages), et leur publication automatique sur les réseaux. Ces usages posent des questions importantes pour la qualité du débat public, la confiance dans l'information, et la lutte contre la désinformation.

La **diversité linguistique et culturelle** est également un sujet d'attention. Les modèles d'IA sont souvent entraînés principalement sur des contenus en anglais ou issus de cultures dominantes. Cela peut entraîner une sous-représentation de certaines langues, expressions culturelles ou visions du monde. La favorisation des corpus majoritaires, risque de marginaliser des langues moins parlées ou de formater les réponses selon des référentiels culturels partiels. Elle peut également réduire la diversité culturelle des contenus produits et partagés.

Enfin, en s'entraînant sur d'immenses corpus issus d'Internet, les modèles des IA génératives peuvent reproduire, transformer ou imiter des œuvres existantes sans que les créateurs soient identifiés, reconnus ou rémunérés. Cette situation remet en question les mécanismes traditionnels de la **propriété intellectuelle**, notamment dans les domaines de création culturelle.

Exemples de savoirs et savoir-faire :

- Citer des exemples bien connus de cas où l'utilisation de système d'IA a abouti à des discriminations des individus (**novice**)
- Décrire le principe de reproduction de biais par un système d'IA (**indépendant**)
- Expliquer en quoi les systèmes de recommandation peuvent propulser de fausses informations (**indépendant**)
- Comparer des IA génératives sur des questions susceptibles de provoquer des résultats avec des biais culturels et linguistiques (**avancé**)

Retrouvez-nous sur :

pix.fr 🔍

Contactez-nous :

pix.fr/support 🔍



pix