

NORME
INTERNATIONALE

ISO/IEC
22989

Première édition
2022-07

**Technologies de l'information —
Intelligence artificielle — Concepts et
terminologie relatifs à l'intelligence
artificielle**

*Information technology — Artificial intelligence — Artificial
intelligence concepts and terminology*



Numéro de référence
ISO/IEC 22989:2022(F)



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO/IEC 2022

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Genève
Tél.: +41 22 749 01 11
Fax: +41 22 749 09 47
E-mail: copyright@iso.org
Web: www.iso.org

Publié en Suisse

Sommaire

	Page
Avant-propos	vi
Introduction	vii
1 Domaine d'application	1
2 Références normatives	1
3 Termes et définitions	1
3.1 Termes relatifs à l'IA	1
3.2 Termes relatifs aux données	6
3.3 Termes relatifs à l'apprentissage machine	8
3.4 Termes relatifs aux réseaux neuronaux	10
3.5 Termes relatifs à la crédibilité	12
3.6 Termes relatifs au traitement du langage naturel	14
3.7 Termes relatifs à la vision artificielle	17
4 Abréviations	18
5 Concepts de l'IA	18
5.1 Généralités	18
5.2 De l'IA forte et faible à l'IA générale et étroite	19
5.3 Agent	19
5.4 Connaissances	20
5.5 Cognition et informatique cognitive	20
5.6 Informatique sémantique	20
5.7 Calcul souple	21
5.8 Algorithmes génétiques	21
5.9 Approches symbolique et sous-symbolique pour l'IA	21
5.10 Données	21
5.11 Concepts liés à l'apprentissage machine	23
5.11.1 Apprentissage supervisé	23
5.11.2 Apprentissage non supervisé	23
5.11.3 Apprentissage semi-supervisé	23
5.11.4 Apprentissage par renforcement	24
5.11.5 Apprentissage par transfert	24
5.11.6 Données d'entraînement	24
5.11.7 Modèle entraîné	24
5.11.8 Données de validation et d'essai	24
5.11.9 Ré-entraînement	25
5.12 Exemples d'algorithmes d'apprentissage machine	26
5.12.1 Réseaux neuronaux	26
5.12.2 Réseaux bayésiens	27
5.12.3 Arbres de décision	28
5.12.4 Machine à vecteurs de support (SVM)	28
5.13 Autonomie, hétéronomie et automatisation	28
5.14 Internet des objets et systèmes cyber-physiques	30
5.14.1 Généralités	30
5.14.2 Internet des objets	30
5.14.3 Systèmes cyber-physiques	30
5.15 Crédibilité	31
5.15.1 Généralités	31
5.15.2 Robustesse de l'IA	31
5.15.3 Fiabilité de l'IA	32
5.15.4 Résilience de l'IA	32
5.15.5 Contrôlabilité de l'IA	32
5.15.6 Explicabilité de l'IA	33
5.15.7 Prévisibilité de l'IA	33

5.15.8	Transparence de l'IA	33
5.15.9	Biais et équité de l'IA	34
5.16	Vérification et validation de l'IA	34
5.17	Questions juridictionnelles	35
5.18	Impact sociétal	35
5.19	Rôles des parties prenantes de l'IA	36
5.19.1	Généralités	36
5.19.2	Fournisseur d'IA	36
5.19.3	Producteur d'IA	37
5.19.4	Client d'IA	37
5.19.5	Partenaire d'IA	37
5.19.6	Sujet d'IA	38
5.19.7	Autorités compétentes	38
6	Cycle de vie de système d'IA	39
6.1	Modèle de cycle de vie de système d'IA	39
6.2	Processus et étapes du cycle de vie d'un système d'IA	41
6.2.1	Généralités	41
6.2.2	Initialisation	41
6.2.3	Conception et développement	43
6.2.4	Vérification et validation	43
6.2.5	Déploiement	43
6.2.6	Exploitation et suivi	44
6.2.7	Validation continue	44
6.2.8	Réévaluation	44
6.2.9	Mise hors service	45
7	Aperçu fonctionnel du système d'IA	45
7.1	Généralités	45
7.2	Données et informations	46
7.3	Connaissance et apprentissage	46
7.4	Des prédictions aux actions	47
7.4.1	Généralités	47
7.4.2	Prédiction	48
7.4.3	Décision	48
7.4.4	Action	48
8	Écosystème de l'IA	48
8.1	Généralités	48
8.2	Systèmes d'IA	50
8.3	Fonction d'IA	50
8.4	Apprentissage machine	50
8.4.1	Généralités	50
8.5	Ingénierie	51
8.5.1	Généralités	51
8.5.2	Systèmes experts	51
8.5.3	Programmation logique	51
8.6	Sources de données et de mégadonnées — informatique en nuage et en périphérie	51
8.6.1	Sources de données et de mégadonnées	51
8.6.2	Informatique en nuage et en périphérie	53
8.7	Réserves de ressources	55
8.7.1	Généralités	55
8.7.2	Circuit intégré spécifique	56
9	Domaines de l'IA	56
9.1	Vision artificielle et reconnaissance d'image	56
9.2	Traitement du langage naturel	57
9.2.1	Généralités	57
9.2.2	Composants de traitement du langage naturel	57
9.3	Exploration de données	60

9.4	Planification.....	60
10	Applications des systèmes d'IA	60
10.1	Généralités	60
10.2	Détection de la fraude	61
10.3	Véhicules automatisés.....	61
10.4	Maintenance prédictive.....	62
Annexe A (informative) Correspondance entre le cycle de vie d'un système d'IA et la définition du cycle de vie d'un système d'IA de l'OCDE.....		63
Bibliographie.....		65

Avant-propos

L'ISO (Organisation internationale de normalisation) et l'IEC (Commission électrotechnique internationale) forment le système spécialisé de la normalisation mondiale. Les organismes nationaux membres de l'ISO ou de l'IEC participent au développement de Normes Internationales par l'intermédiaire des comités techniques créés par l'organisation concernée afin de s'occuper des domaines particuliers de l'activité technique. Les comités techniques de l'ISO et de l'IEC collaborent dans des domaines d'intérêt commun. D'autres organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO et l'IEC participent également aux travaux.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier, de prendre note des différents critères d'approbation requis pour les différents types de documents. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives ou www.iec.ch/members_experts/refdocs).

L'attention est attirée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO et l'IEC ne sauraient être tenues pour responsables de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets) ou dans la liste des déclarations de brevets reçues par l'IEC (voir patents.iec.ch).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la nature volontaire des normes, la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir www.iso.org/iso/fr/avant-propos. Pour l'IEC, voir www.iec.ch/understanding-standards.

Le présent document a été élaboré par le comité technique mixte ISO/IEC JTC 1, *Technologies de l'information*, sous-comité SC 42, *Intelligence artificielle*.

Il convient que l'utilisateur adresse tout retour d'information ou toute question concernant le présent document à l'organisme national de normalisation de son pays. Une liste exhaustive desdits organismes se trouve aux adresses www.iso.org/fr/members.html et www.iec.ch/national-committees.

Introduction

Les progrès réalisés en matière de capacité de calcul, la réduction des coûts de calcul, la disponibilité de grandes quantités de données provenant de nombreuses sources, les programmes d'apprentissage en ligne peu coûteux et les algorithmes capables d'atteindre ou de dépasser les performances humaines dans des tâches particulières en termes de rapidité et d'exactitude ont permis des applications pratiques de l'IA, qui en font une branche de plus en plus importante des technologies de l'information.

L'IA est un domaine fortement interdisciplinaire qui s'appuie largement sur l'informatique, la science des données, les sciences naturelles, les sciences humaines, les mathématiques, les sciences sociales et autres. Des termes tels que «intelligent», «intelligence», «compréhension», «connaissances», «apprentissage», «décisions», «compétences», etc., sont utilisés tout au long du présent document. Toutefois, l'intention n'est pas d'anthropomorphiser les systèmes d'IA, mais de décrire le fait que certains systèmes d'IA peuvent simuler de manière rudimentaire ces caractéristiques.

La technologie de l'IA comporte de nombreux domaines. Ces domaines sont étroitement liés et se développent rapidement, c'est pourquoi il est difficile de conjuguer la pertinence de tous les domaines techniques dans un même espace. La recherche sur l'IA inclut des aspects tels que «l'apprentissage, la reconnaissance et la prédiction», «l'inférence, les connaissances et le langage» et «la découverte, la recherche et la création». La recherche couvre également les interdépendances entre ces aspects^[23].

Le concept de l'IA en tant que flux de processus d'entrée et de sortie est partagé par de nombreux chercheurs en IA, et les recherches relatives à chaque étape de ce processus sont en cours. Des concepts et une terminologie normalisés sont nécessaires aux parties prenantes de cette technologie afin d'être mieux comprises et adoptées par un public plus large. En outre, les concepts et les catégories de l'IA permettent de comparer et de classer les différentes solutions en ce qui concerne les propriétés telles que la crédibilité, la robustesse, la résilience, la fiabilité, l'exactitude, la sûreté, la sécurité et la protection de la vie privée. Cela permet aux parties prenantes de choisir des solutions appropriées pour leurs applications et de comparer la qualité des solutions disponibles sur le marché.

Étant donné que le présent document fournit une définition du terme «IA» au sens d'une discipline uniquement, le contexte de son utilisation peut être décrit comme suit: l'IA est un domaine technique et scientifique consacré aux systèmes techniques qui génèrent des sorties telles que du contenu, des prévisions, des recommandations ou des décisions pour un ensemble donné d'objectifs définis par l'homme.

Le présent document fournit des concepts et une terminologie normalisés afin de contribuer à une meilleure compréhension de la technologie de l'IA et à son utilisation par un plus large éventail de parties prenantes. Il est destiné à un large public, qui réunit experts et néophytes. La lecture de certains articles ou paragraphes spécifiques peut toutefois être plus facile pour qui possède des connaissances plus approfondies dans le domaine de l'informatique. Ces préoccupations sont décrites principalement aux [Paragraphes 5.10 et 5.11](#) et à [l'Article 8](#), qui sont de nature plus technique que le reste du document.

Customer: Xavier Quesnot - Company: GETCAAS - Licence: 1 single user.

Order No.: WS-2025-636418 - IMPORTANT: This file is copyright of IEC, Geneva, Switzerland.

All rights reserved - This file is subject to a licence agreement - Enquiries to Email: sales@iec.ch - Tel.: +41 22 919 02 11.

Technologies de l'information — Intelligence artificielle — Concepts et terminologie relatifs à l'intelligence artificielle

1 Domaine d'application

Le présent document établit la terminologie relative à l'IA et décrit les concepts dans le domaine de l'IA.

Le présent document peut être utilisé dans le cadre de l'élaboration d'autres normes et à l'appui de communications entre parties intéressées ou parties prenantes diverses.

Le présent document est applicable à tous les types d'organismes (par exemple: les entreprises commerciales, les organismes publics, les organismes à but non lucratif).

2 Références normatives

Le présent document ne contient aucune référence normative.

3 Termes et définitions

Pour les besoins du présent document, les termes et les définitions suivants s'appliquent.

L'ISO et l'IEC tiennent à jour des bases de données terminologiques destinées à être utilisées en normalisation, consultables aux adresses suivantes:

- ISO Online browsing platform: disponible à l'adresse <https://www.iso.org/obp>
- IEC Electropedia: disponible à l'adresse <https://www.electropedia.org/>

3.1 Termes relatifs à l'IA

3.1.1

agent intelligent

entité automatisée (3.1.7) qui détecte son environnement, y réagit, et effectue des actions pour atteindre ses objectifs

3.1.2

composant intelligent

élément fonctionnel qui compose un *système d'IA* (3.1.4)

3.1.3

intelligence artificielle

IA

<discipline> recherche et développement de mécanismes et d'applications de *systèmes d'IA* (3.1.4)

Note 1 à l'article: La recherche et le développement peuvent s'opérer dans un grand nombre de domaines tels que l'informatique, la science des données, les sciences humaines, les mathématiques et les sciences naturelles.

3.1.4

système d'intelligence artificielle

système d'IA

système technique qui génère des sorties telles que du contenu, des prévisions, des recommandations ou des décisions pour un ensemble donné d'objectifs définis par l'homme

Note 1 à l'article: Le système technique peut utiliser diverses techniques et approches liées à l'*intelligence artificielle* (3.1.3) pour élaborer un *modèle* (3.1.23) afin de représenter des données, des *connaissances* (3.1.21), des processus, etc., qui peuvent être utilisés pour réaliser des *tâches* (3.1.35).

Note 2 à l'article: Les systèmes d'IA sont conçus pour fonctionner à des niveaux variables d'*automatisation* (3.1.7).

3.1.5

autonomie

autonome

caractéristique d'un système qui est capable de modifier son domaine d'utilisation ou son objectif prévu sans intervention, contrôle ou supervision externe

3.1.6

circuit intégré spécifique

ASIC

circuit intégré personnalisé pour une utilisation particulière

[SOURCE: ISO/IEC IEEE 24765:2017, 3.193, modifié — L'acronyme a été déplacé sur une ligne séparée.]

3.1.7

automatique

automatisation

automatisé

qualifie un processus ou un système qui, dans des conditions déterminées, se déroule ou fonctionne sans intervention humaine

[SOURCE: ISO/IEC 2382:2015, 2121282, modifié — Dans la définition, «un processus ou un appareil» a été remplacé par «un processus ou un système» et les termes privilégiés «automatisé et automatisation» ont été ajoutés.]

3.1.8

informatique cognitive

catégorie de *systèmes d'IA* (3.1.4) qui permet aux personnes et aux machines d'interagir de façon plus naturelle

Note 1 à l'article: Les tâches d'informatique cognitive sont associées à l'*apprentissage machine* (3.3.5), au traitement de la parole, au *traitement du langage naturel* (3.6.9), à la *vision artificielle* (3.7.1) et aux interfaces homme-machine.

3.1.9

apprentissage continu

apprentissage permanent

apprentissage tout au long de la vie

entraînement progressif d'un *système d'IA* (3.1.4) qui a lieu de façon continue pendant la phase d'exploitation du cycle de vie du système d'IA

3.1.10

connexionnisme

paradigme connexionniste

modèle connexionniste

approche connexionniste

forme de modélisation cognitive qui utilise un réseau d'unités interconnectées qui sont généralement des unités de calcul simples

3.1.11**exploration de données**

processus informatique qui extrait des tendances en analysant des données quantitatives selon des perspectives et dimensions différentes, les classe par catégories et en récapitule les relations et impacts potentiels

[SOURCE: ISO 16439:2014, 3.13.]

3.1.12**connaissance déclarative**

connaissance représentée par des faits, des règles ou des théorèmes

Note 1 à l'article: Normalement, la connaissance déclarative ne peut être traitée qu'après avoir été traduite en *connaissance procédurale* ([3.1.28](#)).

[SOURCE: ISO/IEC 2382-28:1995, 28.02.22, modifié — Suppression de la virgule après «règles» dans la version anglaise de la définition.]

3.1.13**système expert**

système d'IA ([3.1.4](#)) qui accumule, associe et encapsule des *connaissances* ([3.1.21](#)) fournies par un ou plusieurs expert(s) humain(s) dans un domaine spécifique afin de déduire des solutions à des problèmes

3.1.14**IA générale****AGI**

intelligence artificielle générale

type de *système d'IA* ([3.1.4](#)) qui traite un large éventail de *tâches* ([3.1.35](#)) en affichant un niveau de performances satisfaisant

Note 1 à l'article: Par rapport à une *IA étroite* ([3.1.24](#)).

Note 2 à l'article: L'AGI est souvent utilisée dans un sens plus large, c'est-à-dire des systèmes qui peuvent non seulement effectuer un large éventail de tâches, mais aussi toutes les tâches qu'un humain peut effectuer.

3.1.15**algorithme génétique****GA**

algorithme qui simule la sélection naturelle en créant et en faisant évoluer une population d'individus (solutions) pour les problèmes d'optimisation

3.1.16**héteronomie****héteronome**

caractéristique d'un système fonctionnant sous la contrainte d'une intervention, d'un contrôle ou d'une supervision externes

3.1.17**inférence**

raisonnement qui consiste à dériver une conclusion d'une prémissse connue

Note 1 à l'article: En IA, une prémissse est un fait, une règle, un modèle, une caractéristique ou des données brutes.

Note 2 à l'article: Le terme «inférence» désigne à la fois le processus et son résultat.

[SOURCE: ISO/IEC 2382:2015, 2123830, modifié – ajout de «un modèle, une caractéristique ou des données brutes». La «Note 4 à l'article: 28.03.01 (2382)» a été supprimée. La «Note 3 à l'article: inférence: terme et définition normalisés par l'ISO/IEC 2382-28:1995» a été supprimée.]

3.1.18

internet des objets

IdO

infrastructure composée d'entités, de personnes, de systèmes et de ressources informationnelles interconnectés, ainsi que de services qui traitent les informations provenant du monde physique et du monde virtuel et y réagissent

[SOURCE: ISO/IEC 20924:2021, 3.2.4, modifié – l'acronyme a été déplacé sur une ligne distincte.]

3.1.19

dispositif de l'IdO

entité d'un *système de l'IdO* ([3.1.20](#)) qui interagit et communique avec le monde physique par la détection ou l'actionnement

Note 1 à l'article: Un dispositif de l'IdO peut être un capteur ou un actionneur.

[SOURCE: ISO/IEC 20924:2021, 3.2.6]

3.1.20

système de l'IdO

système fournissant les fonctionnalités de l'*IdO* ([3.1.18](#))

Note 1 à l'article: Un système de l'IdO peut inclure, sans que cela s'y limite, des dispositifs de l'IdO, des passerelles de l'IdO, des capteurs et des actionneurs.

[SOURCE: ISO/IEC 20924:2021, 3.2.9]

3.1.21

connaissances

<intelligence artificielle> informations résumées relatives à des objets, des événements, des concepts ou des règles, leurs relations et leurs propriétés, organisées en vue d'une utilisation systématique orientée vers un but

Note 1 à l'article: Dans le domaine de l'IA, le terme «connaissances» n'implique pas de capacité cognitive, contrairement à l'usage qui en est fait dans d'autres domaines. Plus particulièrement, les connaissances n'impliquent pas l'action cognitive de compréhension.

Note 2 à l'article: Les informations peuvent exister sous une forme numérique ou symbolique.

Note 3 à l'article: Les informations sont des données qui ont été contextualisées, afin d'être interprétables. Les données sont créées par abstraction ou mesure par rapport au monde réel.

3.1.22

cycle de vie

évolution d'un système, d'un produit, d'un service, d'un projet ou d'une autre entité créée par l'homme, de la conception jusqu'à la mise hors service

[SOURCE: ISO/IEC IEEE 15288:2015, 4.1.23]

3.1.23

modèle

représentation physique, mathématique ou logique d'un système, d'une entité, d'un phénomène, d'un processus ou de données

[SOURCE: ISO/IEC 18023-1:2006, 3.1.11, modifié – Suppression de la virgule après «mathématique» dans la version anglaise de la définition. Ajout de «ou de données» à la fin.]

3.1.24

IA étroite

type de *système d'IA* ([3.1.4](#)) qui se concentre sur des *tâches* ([3.1.35](#)) définies pour traiter un problème spécifique

Note 1 à l'article: Par rapport à une *IA générale* ([3.1.14](#)).

3.1.25
performances
 résultat mesurable

Note 1 à l'article: Les performances peuvent se rapporter à des résultats quantitatifs ou qualitatifs.

Note 2 à l'article: Les performances peuvent concerner la gestion d'activités, de processus, de produits (y compris de services), de systèmes ou d'organismes.

3.1.26
planification
<intelligence artificielle> processus informatiques qui composent un flux de travail à partir d'un ensemble d'actions, en vue d'atteindre un objectif spécifié

Note 1 à l'article: Le terme «planification» utilisé dans le cycle de vie de l'IA ou dans les normes de gestion de l'IA peut également désigner des actions effectuées par des êtres humains.

3.1.27
prédition
sortie principale d'un *système d'IA* ([3.1.4](#)) lorsque lui sont fournies des *données d'entrée* ([3.2.9](#)) ou des informations

Note 1 à l'article: Les prédictions peuvent être suivies de sorties supplémentaires, telles que des recommandations, des décisions et des actions.

Note 2 à l'article: Le terme «prédition» ne se rapporte pas nécessairement au fait de prédire un événement futur.

Note 3 à l'article: Les prédictions peuvent se référer à différents types d'analyse ou de production de données appliquées à de nouvelles données ou à des données historiques (y compris la traduction de textes, la création d'images de synthèse ou le diagnostic d'une panne d'électricité antérieure).

3.1.28
connaissance procédurale
connaissance qui indique de façon explicite la suite d'actions à entreprendre pour résoudre un problème ou atteindre un but

[SOURCE: ISO/IEC 2382-28:1995, 28.02.23]

3.1.29
robot
système d'automatisation doté d'actionneurs qui effectue les *tâches* ([3.1.35](#)) prévues dans le monde physique, en détectant son environnement et en utilisant un système de commande logiciel

Note 1 à l'article: Un robot inclut le système de commande et l'interface d'un système de commande.

Note 2 à l'article: En fonction de son application prévue, un robot est classé comme robot industriel ou comme robot de service.

Note 3 à l'article: Pour effectuer correctement ses *tâches* ([3.1.35](#)), un robot utilise différents types de capteurs pour confirmer son état actuel et percevoir les éléments qui composent l'environnement dans lequel il opère.

3.1.30
robotique
science et pratique de la conception, de la fabrication et de la mise en œuvre de robots

[SOURCE: ISO 8373:2012, 2.16]

3.1.31
informatique sémantique
domaine de l'informatique qui vise à identifier les significations du contenu informatique et les intentions de l'utilisateur et à les exprimer sous une forme exploitable par machine

3.1.32

calcul souple

domaine de calcul qui tolère et exploite l'imprécision, l'incertitude et la vérité partielle pour rendre la résolution de problèmes plus tractable et robuste

Note 1 à l'article: Le calcul souple englobe diverses techniques telles que la logique floue, l'apprentissage machine et le raisonnement probabiliste.

3.1.33

IA symbolique

IA ([3.1.3](#)) basée sur des techniques et des *modèles* ([3.1.23](#)) qui manipulent des symboles et des structures conformément à des règles définies explicitement afin d'obtenir des inférences

Note 1 à l'article: Par rapport à l'*IA sous-symbolique* ([3.1.34](#)), l'*IA symbolique* produit des résultats déclaratifs, tandis que l'*IA sous-symbolique* se fonde sur des approches statistiques et produit des résultats avec une probabilité d'erreur donnée.

3.1.34

IA sous-symbolique

IA ([3.1.3](#)) basée sur des techniques et des *modèles* ([3.1.23](#)) qui utilisent un encodage implicite des informations, qui peuvent être dérivées de l'expérience ou de données brutes

Note 1 à l'article: À distinguer de l'*IA symbolique* ([3.1.33](#)). Tandis que l'*IA symbolique* produit des résultats déclaratifs, l'*IA sous-symbolique* se fonde sur des approches statistiques et produit des résultats avec une probabilité d'erreur donnée.

3.1.35

tâche

<intelligence artificielle>action requise pour atteindre un but spécifique

Note 1 à l'article: Les actions peuvent être physiques ou cognitives. Pour exemple, le calcul ou la création de *prédictions* ([3.1.27](#)), de traductions, de données ou artéfacts de synthèse ou la navigation à l'intérieur d'un espace physique.

Note 2 à l'article: Les exemples de tâches incluent la classification, la régression, le classement, le regroupement et la réduction de la dimensionnalité.

3.2 Termes relatifs aux données

3.2.1

annotation des données

processus consistant à joindre un ensemble d'informations descriptives à des données sans modifier ces données

Note 1 à l'article: Les informations descriptives peuvent prendre la forme de métadonnées, d'étiquettes et d'ancres.

3.2.2

vérification de la qualité des données

processus dans lequel les données sont examinées afin d'établir leur exhaustivité, leur biais et d'autres facteurs qui ont une incidence sur leur utilité pour un *système d'IA* ([3.1.4](#))

3.2.3

augmentation des données

processus consistant à créer des échantillons de synthèse en modifiant ou en utilisant les données existantes

3.2.4**échantillonnages des données**

processus de sélection d'un sous-ensemble d'échantillons de données destiné à présenter des formes et des tendances similaires à celles de l'*ensemble de données* ([3.2.5](#)) plus étendu analysé

Note 1 à l'article: Dans l'idéal, le sous-ensemble d'échantillons de données sera représentatif de l'*ensemble de données* ([3.2.5](#)) plus étendu.

3.2.5**ensemble de données**

collection de données partageant un même format

EXEMPLE 1 Publications de micro-blogging de juin 2020 associées aux mots-dièse #rugby et #football.

EXEMPLE 2 Macrophotographies de fleurs en 256x256 pixels.

Note 1 à l'article: Les ensembles de données peuvent être utilisés pour valider ou soumettre à l'essai un *modèle* ([3.1.23](#)) d'IA. Dans un contexte d'*apprentissage machine* ([3.3.5](#)), les ensembles de données peuvent également être utilisés pour entraîner un *algorithme d'apprentissage machine* ([3.3.6](#)).

3.2.6**analyse exploratoire des données****EDA**

examen initial des données visant à déterminer leurs caractéristiques importantes et à évaluer leur qualité

Note 1 à l'article: L'EDA peut inclure l'identification des valeurs manquantes, des valeurs aberrantes, de la représentativité pour la tâche considérée – voir *vérification de la qualité des données* ([3.2.2](#)).

3.2.7**vérité terrain**

valeur de la variable cible pour un élément particulier de données d'entrée étiquetées

Note 1 à l'article: Le terme «vérité terrain» ne sous-entend pas que les données d'entrée étiquetées correspondent systématiquement à la valeur réelle des variables cibles.

3.2.8**imputation**

procédure dans le cadre de laquelle des données manquantes sont remplacées par des données estimées ou modélisées

[SOURCE: ISO 20252:2019, 3.45]

3.2.9**données d'entrée**

données pour lesquelles un *système d'IA* ([3.1.4](#)) calcule une inférence ou une sortie prévue

3.2.10**étiquette**

variable cible attribuée à un échantillon

3.2.11**données à caractère personnel****DCP****données personnelles**

toute information qui (a) peut être utilisée pour établir un lien entre les informations et la personne physique à laquelle ces informations se rapportent, ou qui (b) est ou peut être associée directement ou indirectement à une personne physique

Note 1 à l'article: La «personne physique» référencée dans la définition est la personne concernée. Pour déterminer si une personne concernée est identifiable, il convient de tenir compte de tous les moyens pouvant être raisonnablement utilisés par la partie prenante en matière de protection de la vie privée qui détient les données, ou par toute autre partie, afin d'établir le lien entre l'ensemble de DCP et la personne physique.

Note 2 à l'article: La présente définition est incluse afin de définir le terme DCP tel qu'utilisé dans le présent document. Un sous-traitant de DCP d'un nuage public n'est généralement pas en position de savoir explicitement si les informations qu'il traite entrent dans une quelconque catégorie spécifiée à moins que le client de services en nuage ne l'indique en toute transparence.

[SOURCE: ISO/IEC 29100:2011/Amd1:2018, 2.9]

3.2.12

données de production

données acquises lors de la phase d'exploitation d'un *système d'IA* ([3.1.4](#)), pour lesquelles un *système d'IA* ([3.1.4](#)) déployé calcule un résultat prévu ou une *inférence* ([3.1.17](#))

3.2.13

échantillon

élément de données atomique traité en quantité par un *algorithme d'apprentissage machine* ([3.3.6](#)) ou un *système d'IA* ([3.1.4](#))

3.2.14

données d'essai

données d'évaluation

données utilisées pour évaluer les performances d'un *modèle* ([3.1.23](#)) final

Note 1 à l'article: Les données d'essai sont distinctes des *données d'entraînement* ([3.3.16](#)) et des *données de validation* ([3.2.15](#)).

3.2.15

données de validation

données de développement

données utilisées pour comparer les performances de différents *modèles* ([3.1.23](#)) candidats

Note 1 à l'article: Les données de validation sont distinctes des *données d'essai* ([3.2.14](#)) et le sont généralement également des *données d'entraînement* ([3.3.16](#)). Toutefois, dans les cas où les données sont insuffisantes pour permettre une division en trois ensembles d'entraînement, de validation et d'essai, les données sont divisées en seulement deux ensembles: un ensemble d'essais et un ensemble d'entraînement ou de validation. La validation croisée ou l'autoamorçage sont des méthodes courantes pour générer des ensembles d'entraînement et de validation distincts à partir de l'ensemble d'entraînement ou de validation.

Note 2 à l'article: Les données de validation peuvent être utilisées pour affiner les hyperparamètres ou pour valider certains choix algorithmiques, jusqu'à l'effet d'inclusion d'une règle donnée dans un système expert.

3.3 Termes relatifs à l'apprentissage machine

3.3.1

réseau bayésien

modèle ([3.1.23](#)) probabiliste qui utilise une *inférence* ([3.1.17](#)) bayésienne pour les calculs de probabilité en utilisant un graphe orienté acyclique

3.3.2

arbre de décision

modèle ([3.1.23](#)) pour lequel l'*inférence* ([3.1.17](#)) est encodée sous la forme de chemins allant de la racine jusqu'à un nœud terminal dans une structure en arbre

3.3.3

collaboration homme-machine

intégration de l'interaction humaine avec les capacités d'intelligence de la machine

3.3.4**hyperparamètre**

caractéristique d'un *algorithme d'apprentissage machine* ([3.3.6](#)) qui a une incidence sur son processus d'apprentissage

Note 1 à l'article: Les hyperparamètres sont choisis avant l'entraînement et peuvent être utilisés dans les processus pour aider à estimer les paramètres de modèle.

Note 2 à l'article: Les exemples d'hyperparamètres incluent le nombre de couches réseau, la largeur de chaque couche, le type de fonction d'activation, la méthode d'optimisation, le facteur d'apprentissage pour les réseaux neuronaux, le choix de la fonction noyau dans une machine à vecteurs de support, le nombre de feuilles ou la profondeur d'un arbre, le K pour le partitionnement K-moyennes, le nombre maximal d'itérations de l'algorithme espérance-maximisation, le nombre de gaussiennes dans un mélange gaussien.

3.3.5**apprentissage machine****ML**

processus d'optimisation des *paramètres de modèle* ([3.3.8](#)) à l'aide de techniques de calcul, de sorte que le comportement du *modèle* ([3.1.23](#)) reflète les données ou l'expérience

3.3.6**algorithme d'apprentissage machine**

algorithme permettant de déterminer les *paramètres* ([3.3.8](#)) d'un *modèle d'apprentissage machine* ([3.3.7](#)) à partir de données conformément à des critères donnés

EXEMPLE Imaginer la résolution d'une fonction linéaire univariée $y = \theta_0 + \theta_1x$ où y est une sortie ou un résultat, x est une entrée, θ_0 est un point d'intersection (la valeur de y où $x = 0$) et θ_1 est un poids. Dans l'*apprentissage machine* ([3.3.5](#)), le processus de détermination du point d'intersection et des poids pour une fonction linéaire est appelé régression linéaire.

3.3.7**modèle d'apprentissage machine**

construction mathématique qui produit une *inférence* ([3.1.17](#)) ou une *prédiction* ([3.1.27](#)) à partir d'informations ou de données d'entrée

EXEMPLE Si une fonction linéaire univariée ($y = \theta_0 + \theta_1x$) a été entraînée à l'aide d'une régression linéaire, le modèle obtenu peut être $y = 3 + 7x$.

Note 1 à l'article: Un modèle d'apprentissage machine découle d'un entraînement basé sur un *algorithme d'apprentissage machine* ([3.3.6](#)).

3.3.8**paramètre****paramètre de modèle**

variable interne d'un *modèle* ([3.1.23](#)) qui a une incidence sur la manière dont il calcule ses sorties

Note 1 à l'article: Les exemples de paramètres incluent les poids d'un réseau neuronal et les probabilités de transition dans un modèle de Markov.

3.3.9**apprentissage par renforcement****RL**

apprentissage d'une suite optimale d'actions dans le but de maximiser une récompense au travers de l'interaction avec un environnement

3.3.10**ré-entraînement**

mise à jour d'un *modèle entraîné* ([3.3.14](#)) par *entraînement* ([3.3.15](#)) avec de nouvelles *données d'entraînement* ([3.3.16](#))

3.3.11

apprentissage semi-supervisé

apprentissage machine (3.3.5) qui utilise à la fois des données étiquetées et des données non étiquetées durant l'entraînement (3.3.15)

3.3.12

apprentissage supervisé

apprentissage machine (3.3.5) qui utilise uniquement des données étiquetées durant l'entraînement (3.3.15)

3.3.13

machine à vecteurs de support

SVM

algorithme d'apprentissage machine (3.3.6) qui trouve les frontières de décision affichant les marges maximales

Note 1 à l'article: Les vecteurs de support sont des ensembles de points de données qui définissent le positionnement des frontières de décision (hyperplans).

3.3.14

modèle entraîné

résultat de l'entraînement de modèle (3.3.15)

3.3.15

entraînement

entraînement de modèle

processus permettant de déterminer ou d'améliorer les paramètres d'un *modèle d'apprentissage machine (3.3.7)*, sur la base d'un *algorithme d'apprentissage machine (3.2.10)*, à l'aide de *données d'entraînement (3.3.16)*

3.3.16

données d'entraînement

données utilisées pour entraîner un *modèle d'apprentissage machine (3.3.7)*

3.3.17

apprentissage non supervisé

apprentissage machine (3.3.5) qui utilise uniquement des données non étiquetées durant l'entraînement (3.3.15)

3.4 Termes relatifs aux réseaux neuronaux

3.4.1

fonction d'activation

fonction appliquée à la combinaison pondérée de toutes les entrées d'un *neurone (3.4.9)*

Note 1 à l'article: Les fonctions d'activation permettent aux réseaux neuronaux d'apprendre des caractéristiques complexes des données. Elles sont généralement non linéaires.

3.4.2

réseau neuronal convolutif

CNN

réseau neuronal convolutif profond

DCNN

réseau neuronal à propagation avant (3.4.6) qui utilise la convolution (3.4.3) dans au moins une de ses couches

3.4.3

convolution

opération mathématique impliquant un produit scalaire glissant ou une corrélation croisée des données d'entrée

3.4.4**apprentissage profond**

apprentissage par réseau neuronal profond

<intelligence artificielle> approche destinée à créer de riches représentations hiérarchiques par l'*entraînement* ([3.3.15](#)) de réseaux neuronaux ([3.4.8](#)) comportant de nombreuses couches cachées

Note 1 à l'article: L'apprentissage profond est un sous-ensemble du *ML* ([3.3.5](#)).

3.4.5**explosion du gradient**

phénomène d'*entraînement* ([3.3.15](#)) par rétropropagation dans un réseau neuronal où de grands gradients d'erreur s'accumulent et entraînent de très grandes mises à jour des poids, ce qui rend le *modèle* ([3.1.23](#)) instable

3.4.6**réseau neuronal à propagation avant****FFNN**

réseau neuronal ([3.4.8](#)) dans lequel les informations sont transmises de la couche d'entrée à la couche de sortie dans une seule direction

3.4.7**mémoire à long et court terme****LSTM**

type de *réseau de neurones récurrents* ([3.4.10](#)) qui traite des données séquentielles en affichant des performances satisfaisantes pour les dépendances de courte durée comme pour celles de longue durée

3.4.8**réseau neuronal****NN**

réseau de neurones formels

réseau de neurones artificiels

<intelligence artificielle> réseau composé d'une ou de plusieurs couches de *neurones* ([3.4.9](#)) reliées entre elles par des connexions pondérées de poids ajustables, qui reçoit des données d'entrée et produit une sortie

Note 1 à l'article: Les réseaux neuronaux constituent un exemple notable de l'*approche connexionniste* ([3.1.10](#)).

Note 2 à l'article: Bien que la conception des réseaux neuronaux s'inspirait initialement du fonctionnement des neurones biologiques, la plupart des travaux sur les réseaux neuronaux ne suivent plus cette voie.

3.4.9**neurone**

<intelligence artificielle> unité élémentaire de traitement qui prend une ou plusieurs valeurs d'entrée et produit une valeur de sortie en associant les valeurs d'entrée et en appliquant une *fonction d'activation* ([3.4.1](#)) au résultat

Note 1 à l'article: Les exemples de fonctions d'activation non linéaires sont une fonction de seuil, une fonction sigmoïde et une fonction polynomiale.

3.4.10**réseau de neurones récurrents****RNN**

réseau neuronal ([3.4.8](#)) dans lequel les sorties de la couche précédente et de l'étape de traitement précédente sont toutes deux transmises à la couche actuelle

3.5 Termes relatifs à la crédibilité

3.5.1

redevable

qui répond des actions, des décisions et des performances

[SOURCE: ISO/IEC 38500:2015, 2.2]

3.5.2

redevabilité

état lié au fait d'être *redevable* (3.5.1)

Note 1 à l'article: La redevabilité se rapporte à une responsabilité attribuée. La responsabilité peut être fondée sur un règlement ou un accord, ou être attribuée dans le cadre d'une délégation.

Note 2 à l'article: La redevabilité implique qu'une personne ou une entité est redevable de quelque chose vis-à-vis d'une autre personne ou entité, via des moyens particuliers et conformément à des critères particuliers.

[SOURCE: ISO/IEC 38500:2015, 2.3, modifié — La Note 2 à l'article a été ajoutée.]

3.5.3

disponibilité

propriété d'être accessible et utilisable à la demande par une entité autorisée

[SOURCE: ISO/IEC 27000:2018, 3.7]

3.5.4

biais

différence systématique de traitement de certains objets, de certaines personnes ou de certains groupes par rapport à d'autres

Note 1 à l'article: Le terme «traitement» désigne tout type d'action, y compris la perception, l'observation, la représentation, la *prédiction* (3.1.27) ou la décision.

[SOURCE: ISO/IEC TR 24027:2021, 3.3.2]

3.5.5

commande; régulation (Syn.)

action délibérée sur (ou dans) un processus, en vue d'atteindre des objectifs définis

[SOURCE: IEC 61800-7-1:2015, 3.2.6]

3.5.6

contrôlabilité

contrôlable

propriété d'un *système d'IA* (3.1.4) qui permet à un humain ou un autre agent externe d'intervenir dans le fonctionnement du système

3.5.7

explicabilité

propriété d'un *système d'IA* (3.1.4) permettant d'exprimer les facteurs importants qui influencent les résultats du *système d'IA* (3.1.4) d'une manière que les humains peuvent comprendre

Note 1 à l'article: Elle a pour objectif de répondre à la question «pourquoi?», sans tenter réellement de défendre l'idée que le plan d'action qui a été entrepris était nécessairement optimal.

3.5.8**prévisibilité**

propriété d'un *système d'IA* ([3.1.4](#)) qui permet aux *parties prenantes* ([3.5.13](#)) de formuler des hypothèses fiables au sujet de la sortie

[SOURCE: ISO/IEC TR 27550:2019, 3.12, «aux personnes, aux propriétaires et aux exploitants de formuler des hypothèses fiables au sujet des DCP et de leur traitement par un système» a été remplacé par «aux parties prenantes de formuler des hypothèses fiables au sujet de la sortie».]

3.5.9**fiabilité**

propriété relative à un comportement et à des résultats prévus et cohérents

[SOURCE: ISO/IEC 27000:2018, 2.55]

3.5.10**résilience**

capacité d'un système à retrouver son état opérationnel rapidement après un incident

3.5.11**risque**

effet de l'incertitude sur les objectifs

Note 1 à l'article: Un effet est un écart par rapport à un attendu. Il peut être positif, négatif ou les deux à la fois et traiter, créer ou entraîner des opportunités et des menaces.

Note 2 à l'article: Les objectifs peuvent avoir différents aspects, être de catégories différentes et peuvent concerner différents niveaux.

Note 3 à l'article: Un risque est généralement exprimé en termes de sources de risque, événements potentiels avec leurs conséquences et leur vraisemblance.

[SOURCE: ISO 31000:2018, 3.1, modifié — Suppression de la virgule après «les deux à la fois» dans la Note 1 à l'article. Suppression de la virgule après «catégories différentes» dans la Note 2 à l'article.]

3.5.12**robustesse**

capacité d'un système à maintenir son niveau de performances dans toutes les situations

3.5.13**partie prenante**

toute personne, groupe ou organisation pouvant affecter, être affecté(e) ou se sentir lui-même/elle-même affecté(e) par une décision ou une activité

[SOURCE: ISO/IEC 38500:2015, 2.24]

3.5.14**transparence**

<organisme> propriété d'un organisme selon laquelle les activités et décisions appropriées sont communiquées aux *parties prenantes* ([3.5.13](#)) pertinentes de manière exhaustive, accessible et compréhensible

Note 1 à l'article: Une communication inappropriée des activités et des décisions peut enfreindre les exigences en matière de sécurité, de protection de la vie privée ou de confidentialité.

3.5.15

transparence

<système> propriété d'un système selon laquelle les informations appropriées relatives au système sont mises à la disposition des *parties prenantes* (3.5.13) pertinentes

Note 1 à l'article: Les informations appropriées pour la transparence du système peuvent inclure des aspects tels que les caractéristiques, les performances, les limites, les composants, les procédures, les mesures, les objectifs de conception, les hypothèses et choix de conception, les sources de données et les protocoles d'étiquetage.

Note 2 à l'article: Une divulgation inappropriée de certains aspects d'un système peut enfreindre les exigences en matière de sécurité, de protection de la vie privée ou de confidentialité.

3.5.16

crédibilité

capacité à satisfaire aux attentes d'une ou de plusieurs *parties prenantes* (3.5.13) de manière vérifiable

Note 1 à l'article: En fonction du contexte ou du secteur, mais aussi du produit ou du service spécifique, des données et de la technologie utilisée, différentes caractéristiques s'appliquent et nécessitent d'être vérifiées pour s'assurer que les attentes des *parties prenantes* (3.5.13) sont satisfaites.

Note 2 à l'article: Les caractéristiques de crédibilité incluent, par exemple, la fiabilité, la disponibilité, la résilience, la sécurité, la protection de la vie privée, la sûreté, la redevabilité, la transparence, l'intégrité, l'authenticité, la qualité et l'utilisabilité.

Note 3 à l'article: La crédibilité est un attribut qui peut être appliqué aux services, aux produits, aux technologies, aux données et aux informations ainsi que, dans le contexte de la gouvernance, aux organismes.

[SOURCE: ISO/IEC TR 24028:2020, 3.42, modifié — «attentes des Parties prenantes» a été remplacé par «attentes des parties prenantes»; la virgule entre «qualité» et «utilisabilité» a été remplacée par «et».]

3.5.17

vérification

confirmation, par des preuves tangibles, que les exigences spécifiées ont été satisfaites

Note 1 à l'article: La vérification garantit uniquement la conformité d'un produit à sa spécification.

[SOURCE: ISO/IEC 27042:2015, 3.21]

3.5.18

validation

confirmation, par la fourniture de preuves objectives, que les exigences relatives à une utilisation prévue ou une application prévue ont été satisfaites

[SOURCE: ISO/IEC 27043:2015, 3.16]

3.6 Termes relatifs au traitement du langage naturel

3.6.1

résumé automatique

tâche (3.1.35) consistant à réduire une partie de contenu ou de texte en *langage naturel* (3.6.7) tout en conservant les informations sémantiques importantes

3.6.2

gestion du dialogue

tâche (3.1.35) consistant à choisir l'action suivante appropriée dans un dialogue sur la base des entrées de l'utilisateur, de l'historique du dialogue et d'autres *connaissances* (3.1.21) contextuelles, afin d'atteindre un objectif souhaité

3.6.3**reconnaissance des émotions**

tâche ([3.1.35](#)) consistant à identifier et à catégoriser par calcul les émotions exprimées dans un texte, un discours, une vidéo, une image ou une combinaison de ces éléments

Note 1 à l'article: Les exemples d'émotions incluent le bonheur, la tristesse, la colère et la joie.

3.6.4**recherche documentaire****IR**

tâche ([3.1.35](#)) consistant à récupérer des documents ou parties de documents pertinents dans un ensemble de données ([3.2.5](#)), généralement basée sur un mot-clé ou sur des requêtes formulées en langage naturel ([3.6.7](#))

3.6.5**traduction automatique****TA**

tâche ([3.1.35](#)) de traduction automatisée d'un texte ou d'un discours d'un *langage naturel* ([3.6.7](#)) dans un autre langage naturel à l'aide d'un système informatique

[SOURCE: ISO 17100:2015, 2.2.2]

3.6.6**reconnaissance d'entités nommées****NER**

tâche ([3.1.35](#)) consistant à reconnaître et à étiqueter les noms dénotationnels d'entités et leurs catégories pour des séquences de mots dans un flux de texte ou de parole

Note 1 à l'article: Le terme «entité» se rapporte à un objet d'intérêt, concret ou abstrait, y compris les associations entre objets.

Note 2 à l'article: «Entité nommée» se rapporte à une entité possédant un nom dénotationnel pour lequel il existe une signification spécifique ou unique.

Note 3 à l'article: Les noms dénotationnels incluent les noms spécifiques des personnes, des lieux, des organismes ou autres noms propres en fonction du domaine ou de l'application.

3.6.7**langage naturel**

langage qui est ou était activement utilisé au sein d'une communauté de personnes et dont les règles sont déduites de l'usage

Note 1 à l'article: Le langage naturel correspond à tout langage parlé par les humains, et qui peut être exprimé sous forme de texte, de discours, de langue des signes, etc.

Note 2 à l'article: Le langage naturel correspond à tout langage parlé par les humains, tel que l'anglais, l'espagnol, l'arabe, le chinois ou le japonais, à ne pas confondre avec les langages de programmation et langages formels, tels que Java, Fortran, C++ ou la logique du premier ordre.

[SOURCE: ISO/IEC 15944-8:2012, 3.82, modifié — «et dont les règles sont principalement déduites de l'usage» remplacé par «et dont les règles sont déduites de l'usage». Suppression de la virgule après «chinois» à la Note 2 à [l'Article 3.6.8](#)]

3.6.8**génération automatique de texte****NLG**

tâche ([3.1.35](#)) consistant à convertir des données contenant de la sémantique en *langage naturel* ([3.6.7](#))

3.6.9

traitement du langage naturel

NLP

<système> traitement des informations basé sur la *compréhension du langage naturel* ([3.6.11](#)) ou la *génération automatique de texte* ([3.6.8](#))

3.6.10

traitement du langage naturel

NLP

<discipline> discipline qui s'intéresse à la manière dont les systèmes acquièrent, traitent et interprètent le *langage naturel* ([3.6.7](#))

3.6.11

compréhension du langage naturel

NLU

compréhension du langage naturel

extraction, par une unité fonctionnelle, des informations contenues dans des textes ou des paroles qui lui sont transmis en *langage naturel* ([3.6.7](#)), et production d'une description de ces textes ou paroles et de ce qu'ils représentent

[SOURCE: ISO/IEC 2382:2015, 2123786, modifié – La Note à l'article a été supprimée, dans la version anglaise le trait d'union de «natural-language» a été supprimé, l'acronyme NLU a été ajouté.]

3.6.12

reconnaissance optique de caractères

OCR

conversion d'images de texte dactylographié, imprimé ou manuscrit en texte encodé à la machine

3.6.13

étiquetage morpho-syntaxique

tâche ([3.1.35](#)) consistant à attribuer une catégorie (par exemple: verbe, nom, adjectif) à un mot en fonction de ses propriétés grammaticales

3.6.14

réponse aux questions

tâche ([3.1.35](#)) consistant à déterminer la réponse la plus appropriée à une question formulée en *langage naturel* ([3.6.7](#))

Note 1 à l'article: Une question peut être ouverte ou être destinée à avoir une réponse spécifique.

3.6.15

extraction de relations

extraction de relations

tâche ([3.1.35](#)) consistant à identifier les relations entre les entités mentionnées dans un texte

3.6.16

analyse de sentiments

tâche ([3.1.35](#)) consistant à identifier et à catégoriser par calcul les opinions exprimées dans un texte, un discours ou une image, afin de déterminer une palette d'émotions, par exemple allant du positif au négatif

Note 1 à l'article: Les exemples de sentiments incluent l'approbation, la désapprobation, l'attitude positive envers, l'attitude négative envers, l'accord et le désaccord.

3.6.17**reconnaissance de la parole**

conversion parole-texte

STT

conversion, par une unité fonctionnelle, d'un signal de parole en une représentation du contenu de la parole

[SOURCE: ISO/IEC 2382:2015, 2120735, modifié — La Note à l'article a été supprimée.]

3.6.18**synthèse de la parole**

conversion texte-parole

TTS

production de parole artificielle

[SOURCE: ISO/IEC 2382: 2015, 2120745]

3.7 Termes relatifs à la vision artificielle**3.7.1****vision artificielle**

capacité d'une unité fonctionnelle à acquérir, traiter et interpréter des données représentant des images ou une vidéo

Note 1 à l'article: La vision artificielle implique l'utilisation de capteurs pour créer une image numérique d'une scène visuelle. Cela peut inclure des images, telles que les images qui captent des longueurs d'onde supérieures à celles de la lumière visible, comme l'imagerie infrarouge.

3.7.2**reconnaissance faciale**

reconnaissance automatique des formes qui compare des images stockées de visages humains avec l'image d'un visage réel, et indique toute correspondance, si elle existe, et toute donnée, si elle existe, identifiant la personne à laquelle le visage appartient

[SOURCE: ISO 5127:2017, 3.1.12.09]

3.7.3**image**

<numérique> contenu graphique destiné à être présenté visuellement

Note 1 à l'article: Cela inclut les graphiques qui sont encodés dans tout format électronique, y compris, sans que cela s'y limite, les formats composés de pixels individuels (par exemple: ceux produits par des programmes de peinture ou par des moyens photographiques) et les formats composés de formules (par exemple: ceux produits sous forme de dessins vectoriels adaptables).

[SOURCE: ISO/IEC 20071-11:2019, 3.2.1]

3.7.4**reconnaissance d'image**processus de classification des images qui classe le(s) objet(s), le(s) forme(s) ou le(s) concept(s) à l'intérieur d'une *image* ([3.7.3](#))

4 Abréviations

API	interface de programmation d'application (application programming interface)
CPS	systèmes cyber-physique (cyber-physical systems)
CPU	unité centrale de traitement (central processing unit)
CRISP-DM	modèle de processus intersectoriel pour l'exploration de données (cross-industry process model for data mining)
DNN	réseau neuronal profond (deep neural network)
DSP	processeur de signal numérique (digital signal processor)
FPGA	réseau de portes programmables sur site (field-programmable gate array)
GPU	processeur graphique (graphics processing unit)
HMM	modèle de Markov caché (hidden Markov model)
KDD	découverte de connaissances à partir de données (knowledge discovery in data)
NPU	unité de traitement de réseau neuronal (neural network processing unit)
OCDE	organisation de coopération et de développement économiques
POS	catégorie grammaticale (part of speech)
TI	technologies de l'information

5 Concepts de l'IA

5.1 Généralités

L'étude interdisciplinaire et le développement de systèmes d'IA visent à construire des systèmes informatiques capables d'effectuer des tâches qui nécessitent normalement une intelligence. Les machines dotées de l'IA sont destinées à percevoir certains environnements et à prendre des mesures qui satisfont leurs besoins.

L'IA utilise des techniques issues de nombreux domaines, tels que l'informatique, les mathématiques, la philosophie, la linguistique, l'économie, la psychologie et les sciences cognitives.

Par rapport à la plupart des systèmes conventionnels non dotés de l'IA, il existe un certain nombre de caractéristiques intéressantes qui sont partagées par tout ou partie des systèmes d'IA.

- a) Interactifs — les entrées des systèmes d'IA sont générées par des capteurs ou par des interactions avec des humains, avec des sorties qui peuvent entraîner la stimulation d'un actionneur ou la fourniture de réponses à des humains ou à des machines. Un exemple peut être la reconnaissance d'objets résultant de la présentation de l'image d'un objet à un système d'IA.
- b) Contextuels — certains systèmes d'IA peuvent s'appuyer sur de multiples sources d'information, dont des informations numériques structurées et non structurées, ainsi que des entrées sensorielles.
- c) Supervision — les systèmes d'IA peuvent fonctionner avec divers degrés de supervision et de contrôle humains, selon l'application. Un exemple serait un véhicule autonome comportant des niveaux variables d'automatisation.

- d) Adaptatifs — certains systèmes d'IA sont conçus pour utiliser des données dynamiques en temps réel et se ré-entrainer afin de mettre à jour leur fonctionnement sur la base des nouvelles données.

5.2 De l'IA forte et faible à l'IA générale et étroite

D'un point de vue philosophique, la faisabilité de machines douées d'intelligence a fait l'objet de débats. Ceux-ci ont débouché sur l'introduction de deux types d'IA différents: l'IA dite «faible» et l'IA dite «forte». Dans l'IA faible, le système peut uniquement traiter des symboles (lettres, nombres, etc.) sans jamais comprendre ce qu'il fait. Dans l'IA forte, le système traite également les symboles, mais il comprend véritablement ce qu'il fait. Les appellations «IA faible» et «IA forte» sont principalement importantes pour les philosophes, mais sont hors sujet pour les chercheurs et les utilisateurs d'IA.

À la suite de ces débats sont apparus les qualificatifs d'«IA étroite » et d'«IA générale», qui sont plus adaptés au domaine de l'IA. Un système d'«IA étroite» est capable de résoudre des tâches définies en vue de traiter un problème spécifique (éventuellement bien mieux que ne le feraient des humains). Un système d'«IA générale» traite un large éventail de tâches en affichant un niveau de performances satisfaisant. Les systèmes d'IA actuels sont considérés comme étant «étroits». Le fait que les systèmes d'IA «générale» soient techniquement possibles à l'avenir reste à établir.

5.3 Agent

Il est possible d'envisager les systèmes d'IA depuis la perspective du paradigme de l'agent étant donné que certaines applications de l'IA visent à simuler l'intelligence humaine et le comportement humain. Définie en tant que discipline technique, l'IA peut être envisagée comme le domaine qui essaie de construire des agents artificiels affichant un comportement rationnel. Le paradigme de l'agent établit une ligne claire entre l'agent et l'environnement dans lequel il évolue. Le paradigme de l'agent est représenté à la [Figure 1](#).

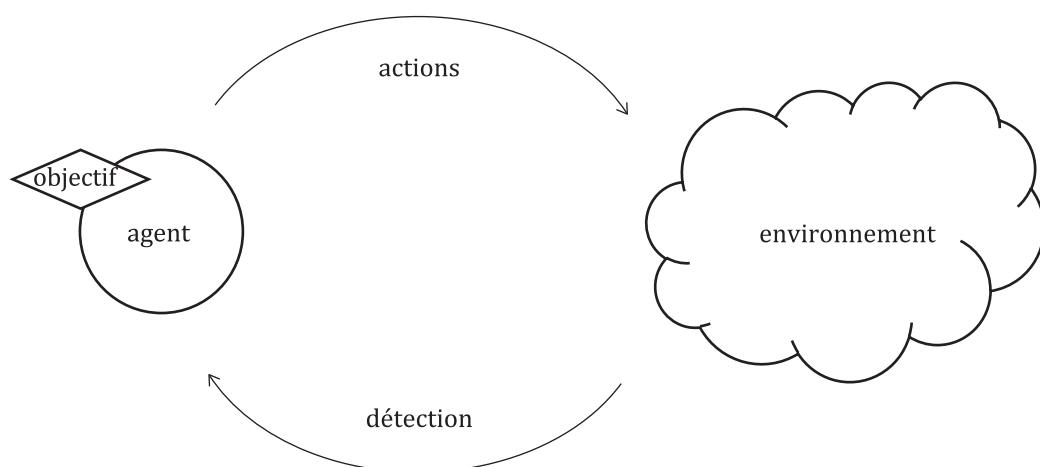


Figure 1 — Paradigme de l'agent

Un agent intelligent interagit avec son environnement via des capteurs et des actionneurs, et effectue des actions qui maximisent ses chances d'atteindre ses objectifs.

Les environnements possèdent différentes caractéristiques en fonction de la tâche entreprise, caractéristiques qui ont une incidence sur le niveau de difficulté de résolution des problèmes.

Dans le présent paradigme, plusieurs types d'agents intelligents peuvent être définis, en fonction de leur architecture^[29]:

- agents réflexes, qui s'appuient uniquement sur la situation actuelle pour choisir une action;
- agents basés sur un modèle, qui dépendent d'un modèle de leur environnement qui leur permet de prendre en compte les résultats de leurs actions disponibles;

- agents basés sur le but ou sur l'utilité, qui dépendent une fonction d'utilité interne qui leur permet de choisir les actions qui atteignent des objectifs, et parmi les objectifs, de rechercher ceux qui sont les plus souhaitables;
- agents apprenants, qui peuvent recueillir des informations au sujet de leur environnement et apprendre à améliorer leurs performances.

Plusieurs architectures complexes et de hauts niveaux basées sur différentes théories ont été développées pour mettre en œuvre les agents.

5.4 Connaissances

La signification de «connaissances» spécifique à l'IA justifie une discussion plus approfondie, en raison de la prévalence de ce concept dans le document et sur le terrain.

Bien que dans d'autres domaines ce terme puisse être associé aux capacités cognitives, dans le contexte de l'IA il s'agit d'un terme purement technique qui se rapporte au contenu, pas aux capacités. Le concept de connaissances fait partie de la hiérarchie données-informations-connaissances, sur la base de laquelle les données peuvent être utilisées pour produire des informations, et les informations peuvent être utilisées pour produire des connaissances. Dans le contexte de l'IA, il s'agit de processus purement techniques et non cognitifs.

Les connaissances se distinguent des informations dans le sens où les informations sont observées par le système, tandis que les connaissances sont ce que le système retient de ces observations. Les connaissances sont structurées et organisées; elles font abstraction des spécificités des observations individuelles. En fonction de l'objet, les mêmes informations peuvent produire des connaissances différentes.

Les connaissances diffèrent de leur représentation dans le sens où les mêmes connaissances peuvent avoir différentes représentations: elles peuvent apparaître sous différentes formes concrètes, chacune avec ses avantages et inconvénients, mais elles ont toutes la même signification.

Ces distinctions ont un impact technique, car certaines approches, méthodes et autres sujets d'étude liés à l'IA dépendent entièrement de la capacité à produire des connaissances différentes pour les mêmes informations, ou des représentations différentes pour les mêmes connaissances.

5.5 Cognition et informatique cognitive

La cognition comprend l'acquisition et le traitement de connaissances par le raisonnement, l'expérience exclusive ou partagée, l'apprentissage et la perception. Elle englobe des concepts tels que l'attention, la formation de connaissances, la mémoire, le jugement et l'évaluation, le raisonnement et le calcul, la résolution de problèmes et la prise de décisions, la compréhension et la production de langage.

L'informatique cognitive compte parmi les sous-disciplines de l'IA^[27]. Elle vise à mettre en œuvre la cognition en utilisant des capacités telles que l'identification de formes à partir du traitement d'énormes quantités d'informations. Elle permet aux personnes d'interagir avec les machines de façon plus naturelle. Les tâches d'informatique cognitive sont associées à l'apprentissage machine, au traitement de la parole, au traitement du langage naturel, à la vision artificielle et aux interfaces homme-machine.

5.6 Informatique sémantique

L'informatique sémantique traite de la mise en correspondance de la sémantique du contenu informatique et des intentions humaines. Elle fournit des représentations qui permettent de décrire les informations et utilise ces représentations pour récupérer, gérer, manipuler et créer du contenu (tel que du texte, des vidéos, des fichiers audio, des processus, des fonctions, des dispositifs et des réseaux). La description sémantique du contenu permet de réduire l'incertitude des processus cognitifs et un raisonnement logique au sujet des informations. Cela permet alors l'enrichissement, la coordination, la récapitulation et la comparaison des informations. L'informatique sémantique est par conséquent une approche qui associe informations préalables et apprentissage.

5.7 Calcul souple

Le calcul souple est une méthodologie qui associe différentes techniques qui peuvent tolérer l'imprécision, l'incertitude et la vérité partielle pour résoudre des problèmes complexes. Les méthodes de calcul conventionnelles sont généralement appliquées pour trouver des solutions précises et rigoureuses à des problèmes. Toutefois, ces solutions peuvent être inadaptées ou produire une complexité extrême. Le calcul souple est construit sur la compréhension du caractère souvent imprécis et incertain du monde réel. De ce fait, chercher à établir des solutions précises aux problèmes du monde réel peut souvent avoir pour corollaire des coûts et une certaine complexité. Par conséquent, le calcul souple cherche à tirer parti de la tolérance en matière d'imprécision, d'incertitude et de vérité partielle pour obtenir des solutions tractables, robustes et à faible coût^[24]. Les systèmes flous, les algorithmes évolutionnistes, l'intelligence distribuée et les systèmes à réseau neuronal sont des exemples de techniques de calcul souple.

5.8 Algorithmes génétiques

Les algorithmes génétiques simulent la sélection naturelle en créant et en faisant évoluer une population d'individus (solutions) pour les problèmes d'optimisation. La création de nouvelles solutions basée sur une population initiale s'inspire des mutations génétiques. Le chromosome (série de «gènes») est représenté par une chaîne de 0 et de 1. Une fois qu'une population initiale de chromosomes est générée, la première étape consiste simplement à calculer l'aptitude à l'emploi de chaque chromosome. La valeur de la fonction d'aptitude à l'emploi quantifie l'optimalité d'une solution en la classant par rapport aux autres solutions. Si la solution créée n'est pas optimale, une paire de chromosomes est choisie pour l'échange de parties (croisement) et la création de deux chromosomes descendants. À l'étape suivante, une mutation modifie de façon aléatoire au moins un gène des chromosomes. La population initiale est remplacée par la nouvelle population et une nouvelle itération commence. Les itérations de GA se terminent quand l'un des critères de fin (généralement un nombre prédéfini d'itérations) est satisfait. À la fin, les chromosomes les plus forts sont conservés^[25].

5.9 Approches symbolique et sous-symbolique pour l'IA

Dans la discipline de l'IA, il existe de nombreux points de vue différents avec des paradigmes différents. Il n'y a pas de classification qui permette d'établir une distinction nette entre les différents types d'IA. Il est néanmoins possible de fournir certaines dimensions sur lesquelles les systèmes d'IA peuvent être positionnés.

Depuis la création de l'IA en tant que discipline, deux paradigmes concurrents se sont développés: l'IA symbolique et l'IA sous-symbolique.

L'IA symbolique implique d'encoder les connaissances à l'aide de symboles et de structures et utilise principalement des logiques pour modéliser les processus de raisonnement. Dans ce paradigme, les informations sont encodées explicitement à l'aide d'une représentation formelle, dont la syntaxe est traitable par un ordinateur et dont la sémantique est compréhensible pour les humains.

L'autre approche est l'IA sous-symbolique, qui utilise le paradigme connexioniste. Ce paradigme n'est pas fondé sur le raisonnement symbolique; il s'appuie plutôt sur l'encodage implicite des connaissances. Cette représentation implicite des connaissances se fonde principalement sur des approches statistiques du traitement de l'expérience ou des données brutes. Des exemples de ce type de système d'IA sont les différents systèmes d'apprentissage machine, y compris les différentes formes de réseaux neuronaux profonds.

En règle générale, les systèmes d'IA modernes contiennent à la fois des éléments d'IA symbolique et d'IA sous-symbolique. Ces systèmes sont appelés IA hybride.

5.10 Données

Les données sont au cœur de nombreux systèmes d'IA. Nombre d'entre eux sont conçus pour traiter les données, et il est souvent nécessaire d'utiliser des données à des fins d'essai. Dans le cas de systèmes d'apprentissage machine, leur cycle de vie dépend intégralement de la disponibilité des données.

Les données peuvent se présenter sous forme structurée (par exemple: les bases de données relationnelles) ou sous forme non structurée (par exemple: courriers électroniques, documents texte, images, enregistrements audio et fichiers). Les données sont un aspect clé des systèmes d'IA et elles subissent des processus, parmi lesquels:

- l'acquisition de données, dans laquelle les données sont obtenues auprès d'une ou de plusieurs sources. Les données peuvent être obtenues au sein de l'organisme ou être amenées de l'extérieur. Il est nécessaire que l'adéquation des données soit évaluée, par exemple afin d'établir si elles présentent un quelconque biais ou si elles sont suffisamment larges pour être représentatives des données d'entrée opérationnelles attendues;
- analyse exploratoire des données, dans laquelle les caractéristiques des données sont examinées afin d'établir les formes, les relations, les tendances et les valeurs aberrantes. Cette analyse peut guider des étapes ultérieures telles que l'entraînement et la vérification;
- annotation des données, dans laquelle les éléments importants des données sont ajoutés sous la forme de métadonnées (par exemple: informations relatives à la provenance des données ou étiquettes destinées à faciliter l'entraînement d'un modèle);
- préparation des données, dans laquelle les données sont mises sous une forme qui peut être utilisée par le système d'IA;
- filtrage, qui correspond à la suppression des données indésirables. Il est nécessaire d'examiner soigneusement les effets du filtrage afin d'éviter l'introduction d'un biais indésirable et d'autres problèmes;
- normalisation, qui désigne l'ajustement des valeurs des données à une échelle commune afin qu'elles soient mathématiquement comparables;
- désidentification ou autres processus, qui peuvent être requis si l'ensemble de données inclut des données à caractère personnel (DCP) ou est associé à des personnes ou des organismes, avant que les données puissent être utilisées par le système d'IA (par exemple: voir l'ISO/IEC 20889);
- vérification de la qualité des données, dans laquelle le contenu des données est examiné afin de vérifier son exhaustivité, d'identifier les biais et autres facteurs ayant une incidence sur son utilité pour le système d'IA. Il est essentiel de vérifier qu'il n'y a pas eu empoisonnement des données afin de s'assurer que les données d'entraînement n'ont pas été contaminées par des données pouvant produire des résultats préjudiciables ou indésirables;
- échantillonnage des données, dans lequel un sous-ensemble représentatif des données est extrait;
- augmentation des données, dans laquelle les données qui sont disponibles en trop petite quantité subissent différents types de transformations dans le but d'agrandir l'ensemble de données;
- étiquetage des données, dans lequel les ensembles de données sont étiquetés, ce qui signifie que des échantillons sont associés à des variables cibles. Des étiquettes sont souvent nécessaires pour les données d'essai et les données de validation. Certaines approches en matière de ML dépendent également de la disponibilité d'étiquettes pour l'entraînement du modèle (voir [5.11.1](#) et [5.11.3](#)).

En fonction du cas d'application et de l'approche utilisée, les données d'un système d'IA peuvent être impliquées de plusieurs façons:

- les données de production sont les données traitées par le système d'IA dans la phase d'exploitation. Les systèmes d'IA n'impliquent pas tous des données de production, selon le cas d'application, mais ceci est indépendant de la conception technique et de l'approche du système d'IA;
- les données d'essai sont les données utilisées pour évaluer les performances du système d'IA avant son déploiement. Elles sont censées être similaires aux données de production, et une évaluation adéquate nécessite que les données d'essai soient distinctes de toutes les données utilisées pendant le développement. Toutes les approches en matière d'IA justifient une évaluation, mais en fonction de la tâche, il n'est pas toujours adéquat d'utiliser des données d'essai;

- les données de validation correspondent aux données utilisées par le développeur pour faire certains choix algorithmiques ou les valider (recherche d'hyperparamètre, conception des règles, etc.). Elles ont différents noms selon le domaine de l'IA, par exemple dans le traitement du langage naturel elles sont généralement appelées données de développement. Il existe des cas où aucune donnée de validation n'est nécessaire;
- les données d'entraînement sont utilisées spécifiquement dans le contexte de l'apprentissage machine: elles servent de matières premières à partir desquelles l'algorithme d'apprentissage machine extrait son modèle pour traiter une tâche donnée.

NOTE 1 Dans les cadres d'évaluation logicielle, la validation est le processus consistant à vérifier que certaines exigences ont été satisfaites. Elle fait partie du processus d'évaluation. Dans un contexte spécifique à l'IA, le terme «validation» est utilisé pour désigner le processus consistant à tirer parti des données pour définir certaines valeurs et propriétés pertinentes pour la conception du système. Elle ne concerne pas l'évaluation du système par rapport à ses exigences et elle intervient avant l'étape d'évaluation.

NOTE 2 Dans les cadres d'évaluation logicielle, le terme «essai» se rapporte à divers processus tels que la recherche de bogues, la réalisation d'essais d'unité et la mesure du temps de calcul. Sa signification dans le domaine de l'IA se rapporte spécifiquement à l'évaluation statistique des performances du système par rapport à un ensemble de données dédié.

5.11 Concepts liés à l'apprentissage machine

5.11.1 Apprentissage supervisé

L'apprentissage supervisé est défini comme étant un «apprentissage machine qui utilise des données étiquetées durant l'entraînement» ([3.3.12](#)). Dans ce cas, les modèles de ML sont entraînés avec des données d'entraînement qui incluent une sortie ou une variable cible connue ou déterminée (l'étiquette). La valeur de la variable cible pour un échantillon donné est également appelée «vérité terrain». Les étiquettes peuvent être de tout type y compris des valeurs catégoriques, binaires ou numériques, ou des objets structurés (par exemple: séquences, images, arbres ou graphiques) selon la tâche. Les étiquettes peuvent faire partie de l'ensemble de données d'origine, mais dans de nombreux cas elles sont déterminées manuellement ou au travers d'autres processus.

L'apprentissage supervisé peut être utilisé pour les tâches de classification et de régression, ainsi que pour les tâches plus complexes relatives à la prédiction structurée.

Pour plus d'informations sur l'apprentissage supervisé, voir l'[ISO/IEC 23053](#).

5.11.2 Apprentissage non supervisé

L'apprentissage non supervisé est défini comme étant un «apprentissage machine qui utilise des données non étiquetées durant l'entraînement» ([3.3.17](#)).

L'apprentissage non supervisé peut être utile dans les cas tels que le regroupement où l'objectif de la tâche est de déterminer les points communs entre les échantillons dans les données d'entrée. La réduction de la dimensionnalité d'un ensemble de données d'entraînement est une autre application de l'apprentissage non supervisé dans lequel les caractéristiques les plus pertinentes d'un point de vue statistique sont déterminées indépendamment de toute étiquette.

Pour plus d'informations sur l'apprentissage non supervisé, voir l'[ISO/IEC 23053](#).

5.11.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé est défini comme étant un «apprentissage machine qui utilise à la fois des données étiquetées et des données non étiquetées durant l'entraînement» ([3.3.11](#)). L'apprentissage semi-supervisé est un hybride de l'apprentissage supervisé et non supervisé.

L'apprentissage semi-supervisé est utile dans les cas où l'étiquetage de tous les échantillons d'un grand ensemble de données d'entraînement serait prohibitif en termes de temps ou de coût. Voir l'ISO/IEC 23053 pour plus de détails au sujet de l'apprentissage semi-supervisé.

5.11.4 Apprentissage par renforcement

L'apprentissage par renforcement est le processus d'entraînement d'un ou de plusieurs agents qui interagit (interagissent) avec son (leur) environnement afin d'atteindre un but prédéfini. Dans l'apprentissage par renforcement, un ou plusieurs agent(s) d'apprentissage machine apprend (apprennent) par le biais d'un processus itératif d'essais et d'erreurs. L'objectif de l'agent (des agents) est de trouver la stratégie (c'est-à-dire de construire un modèle) permettant d'obtenir de l'environnement les meilleures récompenses. Pour chaque tentative (réussie ou non), l'environnement fournit une rétroaction indirecte. Le(s) agent(s) ajuste(-nt) alors son (leur) comportement (c'est-à-dire son/leur modèle) sur la base de cette rétroaction. Voir l'ISO/IEC 23053 pour plus d'informations sur l'apprentissage par renforcement.

5.11.5 Apprentissage par transfert

L'apprentissage par transfert désigne une série de méthodes dans lesquelles les données destinées à résoudre un problème sont exploitées afin d'appliquer à un autre problème les connaissances acquises grâce à elles. Par exemple, les informations acquises en identifiant les numéros de maison dans la vue d'une rue peuvent être utilisées pour reconnaître des numéros manuscrits. Voir l'ISO/IEC 23053 pour plus de détails au sujet de l'apprentissage par transfert.

5.11.6 Données d'entraînement

Les données d'entraînement se composent d'échantillons de données utilisés pour entraîner un algorithme d'apprentissage machine. En général, les échantillons de données se rapportent à une thématique particulière et ils peuvent se composer de données structurées ou non structurées. Les échantillons de données peuvent être non étiquetés ou étiquetés.

Dans ce dernier cas, l'étiquette est utilisée pour guider le processus d'entraînement du modèle d'apprentissage machine. Par exemple, lorsque les données d'entrée sont des images et que l'objectif est de déterminer si une image montre un chat, l'étiquette peut être «vrai» pour une image représentant un chat et «faux» pour une image qui ne représente pas un chat. Cela permet au modèle entraîné de représenter une relation statistique entre les attributs d'un échantillon de données d'entraînement et la variable cible.

Le nombre d'échantillons de données dans les données d'entraînement et le choix de caractéristiques appropriées contribuent à l'adéquation du modèle ML obtenu avec la distribution des données ou la variable cible. Il existe toutefois la contrepartie du temps de calcul et des ressources nécessaires pour le calcul si l'ensemble de données est très grand.

5.11.7 Modèle entraîné

Le présent document définit un modèle entraîné comme le résultat de l'entraînement d'un modèle, qui a son tour est défini comme étant le processus visant à déterminer ou à améliorer les paramètres d'un modèle d'apprentissage machine, sur la base d'un algorithme d'apprentissage machine, à l'aide de données d'entraînement. Un modèle d'apprentissage machine est une construction mathématique qui génère une inférence, ou une prédition, sur la base d'informations ou de données d'entrée. Il convient que le modèle entraîné soit utilisable par un système d'IA pour formuler des prédictions sur la base de données de production du domaine d'intérêt. Il existe différents formats normalisés qui permettent de stocker et de transmettre le modèle entraîné en tant que jeu de nombres.

5.11.8 Données de validation et d'essai

Pour évaluer le modèle entraîné, il est courant de diviser les données acquises pour le développement d'un modèle en ensemble de données d'entraînement, de validation et d'essai.

Les données de validation sont utilisées pendant et après l'entraînement afin d'ajuster les hyperparamètres. L'ensemble d'essais est utilisé pour vérifier que le modèle a appris ce qu'il était censé apprendre. Tous deux sont composés de données qui ne sont jamais présentées au modèle pendant l'entraînement. Si des données d'entraînement étaient utilisées à ces fins, le modèle est capable de «se souvenir» de la prédiction correcte sans réellement traiter l'échantillon de données. Afin d'éviter de surestimer les performances du modèle, les données d'essai ne sont pas présentées non plus durant le réglage.

En cas d'utilisation de la validation croisée, les données sont divisées de sorte que chaque échantillon de données est utilisé à la fois pour l'entraînement et pour la validation. Cette approche imite l'utilisation d'un ensemble de données plus grand, ce qui peut améliorer les performances du modèle. Parfois, les données sont insuffisantes pour permettre des ensembles d'entraînement, de validation et d'essai distincts. Dans ce cas, les données sont divisées seulement en deux ensembles, c'est-à-dire 1. données d'entraînement ou de validation, et 2. données d'essai. Des ensembles de données de validation et d'entraînement distincts sont alors générés à partir des données d'entraînement ou de validation, par exemple via l'autoamorçage ou la validation croisée.

5.11.9 Ré-entraînement

5.11.9.1 Généralités

Le ré-entraînement consiste à mettre à jour un modèle entraîné en l'entraînant avec des données d'entraînement différentes. Cela peut s'avérer nécessaire en raison de nombreux facteurs, y compris l'absence d'ensembles de données de grande taille, la dérive des données et la dérive conceptuelle.

Dans la dérive des données, l'exactitude des prédictions du modèle se détériore au fil du temps en raison de modifications des caractéristiques statistiques des données de production (par exemple: la résolution des images a changé, ou une classe est devenue plus fréquente qu'une autre dans les données). Dans ce cas, il est nécessaire de ré-entraîner le modèle avec de nouvelles données d'entraînement qui représentent mieux les données de production actuelles.

Dans la dérive conceptuelle, la frontière de décision se déplace (par exemple: ce qui est légal et ce qui ne l'est pas ont tendance à changer avec l'introduction de nouvelles lois), ce qui dégrade également l'exactitude des prédictions, même si les données n'ont pas changé. Dans le cas de la dérive conceptuelle, il est nécessaire de réétiqueter les variables cibles des données d'entraînement et de ré-entraîner le modèle.

Lors du ré-entraînement d'un modèle existant, une considération spécifique consiste à surmonter ou à réduire au minimum les défis de l'oubli dit catastrophique. De nombreux algorithmes d'apprentissage machine sont doués pour apprendre des tâches uniquement si les données sont toutes présentées en même temps. Étant donné qu'un modèle est entraîné à une tâche particulière, ses paramètres sont adaptés afin de résoudre la tâche. Lorsque de nouvelles données d'entraînement sont introduites, les adaptations basées sur ces nouvelles observations effacent et remplacent les connaissances que le modèle avait préalablement acquises. Pour les réseaux neuronaux, ce phénomène est appelé «oubli catastrophique», et il est considéré comme l'une de leurs limites fondamentales.

5.11.9.2 Apprentissage continu

L'apprentissage continu, également appelé apprentissage permanent ou apprentissage tout au long de la vie, est l'entraînement incrémentiel d'un modèle qui se déroule de façon continue tandis que le système fonctionne en production. Il s'agit d'un cas particulier de ré-entraînement, dans lequel les mises à jour du modèle sont répétées, surviennent à une haute fréquence, et n'impliquent aucune interruption de l'exploitation.

Dans de nombreux systèmes d'IA, le système est entraîné pendant le processus de développement, avant sa mise en production. Cette approche est similaire à celle du développement logiciel standard, où le système est construit et intégralement soumis à essai avant sa mise en production. Le comportement de ces systèmes est évalué durant le processus de vérification et est censé être inchangé durant la phase d'exploitation.

Les systèmes d'IA qui représentent l'apprentissage continu impliquent la mise à jour incrémentielle du modèle dans le système tandis qu'il fonctionne pendant la production. Les données entrées du système pendant l'exploitation sont non seulement analysées pour produire une sortie du système, mais aussi utilisées simultanément pour ajuster le modèle dans le système, dans le but d'améliorer le modèle sur la base des données de production. En fonction de la conception du système d'IA à apprentissage continu, des actions humaines peuvent être nécessaires dans le processus, par exemple l'étiquetage des données, la validation de l'application d'une mise à jour incrémentielle spécifique ou la surveillance des performances du système d'IA.

L'apprentissage continu peut aider à gérer les limites des données d'apprentissage d'origine et la dérive conceptuelle et des données. Toutefois, l'apprentissage continu pose d'importants problèmes lorsqu'il s'agit de garantir que le système d'IA continue de fonctionner correctement à mesure qu'il apprend. La vérification du système en production est nécessaire, tout comme la nécessité de capturer les données de production pour les intégrer à l'ensemble de données d'entraînement au cas où le système d'IA serait mis à jour à un moment ultérieur.

En raison du risque d'oubli catastrophique, l'apprentissage continu implique la capacité à apprendre au fil du temps en intégrant de nouvelles observations faites sur les données actuelles, tout en conservant les connaissances antérieures.

Les caractéristiques de l'apprentissage continu incluent:

- apprentissage au fil du temps dans des environnements dynamiques (idéalement dans le monde ouvert);
- augmentation des connaissances précédemment acquises par l'apprentissage de nouvelles connaissances afin d'améliorer les performances (soit par le biais de nouvelles données, soit par le raisonnement sur les connaissances existantes);
- découverte de nouveaux aspects de la tâche à apprendre et leur apprentissage progressif;
- apprentissage en cours d'exécution ou pendant que le système fonctionne en production.

5.12 Exemples d'algorithmes d'apprentissage machine

5.12.1 Réseaux neuronaux

5.12.1.1 Généralités

Les réseaux neuronaux tentent de simuler une capacité intelligente dans l'observation, l'apprentissage, l'analyse et la prise de décision pour des problèmes complexes. La conception des réseaux neuronaux s'inspire donc de la manière dont les neurones sont connectés dans le cerveau des humains et des animaux. La structure des réseaux neuronaux est composée d'éléments de traitement interconnectés, appelés neurones. Chaque neurone reçoit plusieurs entrées et ne génère qu'une seule sortie. Ils sont organisés en couches, la sortie d'une couche devenant l'entrée de la couche suivante. Chaque connexion se voit attribuer un poids lié à l'importance de l'entrée. Le réseau neuronal «apprend» en s'entraînant avec des entrées connues, en comparant la sortie réelle à la sortie attendue et en utilisant l'erreur pour ajuster les poids. Ainsi, les liens qui produisent des réponses correctes sont renforcés et ceux qui génèrent des réponses incorrectes sont affaiblis.

Le présent document définit l'apprentissage profond comme une approche permettant de créer des représentations hiérarchiques riches grâce à l'entraînement de réseaux neuronaux comportant de nombreuses couches cachées. Ce processus permet au réseau neuronal d'affiner progressivement la sortie finale. L'apprentissage profond peut réduire ou éliminer le besoin d'ingénierie des caractéristiques, car les caractéristiques les plus pertinentes sont identifiées automatiquement. L'apprentissage en profondeur peut nécessiter beaucoup de temps et d'importantes ressources informatiques.

Il existe de nombreuses «architectures» de réseaux neuronaux (fondamentalement, des arrangements de neurones) et il s'agit d'un domaine de recherche actif où de nouvelles architectures de réseaux neuronaux continuent d'être introduites. Les exemples d'architectures de NN incluent:

- réseau neuronal à propagation avant;
- réseau de neurones récurrents;
- réseau neuronal convolutif.

Ces architectures de NN sont décrites aux [paragraphes 5.12.1.2 à 5.12.1.4](#)

NOTE Voir l'ISO/IEC 23053 pour plus d'informations sur les NN.

5.12.1.2 Réseau neuronal à propagation avant

Le FFNN constitue l'architecture de réseau neuronal la plus simple. Elle transmet les informations de l'entrée vers la sortie dans une seule direction. Il n'existe pas de connexion entre les neurones d'une même couche. Deux couches adjacentes peuvent être généralement «entièrement connectées», c'est-à-dire que chaque neurone d'une couche possède une connexion avec chaque neurone de la couche suivante.

5.12.1.3 Réseau de neurones récurrents

5.12.1.3.1 Généralités

Les RNN^[21] gèrent les entrées qui apparaissent dans une séquence ordonnée, c'est-à-dire que l'ordre des entrées dans la séquence compte. Les exemples d'entrées de ce type incluent les séquences dynamiques telles que les flux audio et vidéo, mais aussi les séquences statiques telles que le texte ou même les images uniques. Les RNN sont dotés de noeuds qui à la fois reçoivent les informations d'entrée de la couche précédente mais également prennent en compte leurs propres informations issues d'un passage antérieur. Les RNN ont une propriété à états influencée par l'apprentissage antérieur. Les RNN sont largement utilisés dans la reconnaissance de la parole, la traduction automatique, la prévision de séries chronologiques et la reconnaissance d'image. Voir l'ISO/IEC 23053 pour plus d'informations sur les RNN.

5.12.1.3.2 Réseau à mémoire à long et court terme

Un réseau LSTM est une forme de RNN conçu pour les problèmes qui nécessitent de mémoriser des informations ayant des différences chronologiques à la fois plus longues et plus courtes, ce qui les rend adaptés à l'apprentissage des connexions à long terme. Ils ont été introduits dans le but de résoudre le problème de disparition du gradient dans les RNN associés à une rétropropagation^[22].

Les réseaux LSTM peuvent apprendre des séquences complexes, comme écrire comme Shakespeare ou composer de la musique. Voir l'ISO/IEC 23053 pour plus d'informations sur les LSTM.

5.12.1.4 Réseau neuronal convolutif (CNN)

Un CNN est un réseau neuronal qui inclut au moins une couche de convolution pour filtrer les informations utiles issues des entrées. Les utilisations courantes incluent la reconnaissance d'image, l'étiquetage vidéo et le traitement du langage naturel. Voir l'ISO/IEC 23053 pour plus d'informations sur les CNN.

5.12.2 Réseaux bayésiens

Les réseaux bayésiens sont des modèles graphiques utilisés pour générer des prédictions au sujet des dépendances entre les variables. Ils peuvent être utilisés pour dériver des probabilités pour les causes ou des variables qui peuvent contribuer au résultat. Cette causalité est très utile dans les applications telles que le diagnostic médical. D'autres applications où les réseaux bayésiens sont utiles incluent

l'analyse des données, le traitement des données incomplètes et l'atténuation du surajustement des modèles aux données. Les réseaux bayésiens reposent sur la probabilité bayésienne: la probabilité d'un événement dépend du degré de croyance dans cet événement. De plus amples informations sur les réseaux bayésiens sont disponibles dans^[20] et dans l'ISO/IEC 23053.

5.12.3 Arbres de décision

Les arbres de décision utilisent une structure arborescente de décisions pour coder les résultats possibles. Les algorithmes d'arbres de décision sont largement utilisés pour la classification et la régression. L'arbre est formé de nœuds de décision et de nœuds terminaux. Chaque nœud de décision possède au moins deux branches, tandis que les nœuds terminaux représentent la décision finale ou la classification. En général, les nœuds sont organisés en fonction de la décision qui donne le prédicteur le plus fort. Il est nécessaire de diviser les données d'entrée en plusieurs facteurs afin de déterminer le meilleur résultat. Les arbres de décision sont analogues aux organigrammes où, à chaque nœud de décision, une question peut être posée pour déterminer sur quelle branche poursuivre.

5.12.4 Machine à vecteurs de support (SVM)

La SVM est une méthode d'apprentissage machine largement utilisée pour la classification et la régression. Les SVM classent les échantillons de données dans deux catégories différentes, puis elles affectent les nouveaux exemples de données à une catégorie ou à l'autre. Les SVM sont des algorithmes de classification à distance maximale. Ils définissent un hyperplan pour séparer deux classes au-dessus et au-dessous de celui-ci, en fournissant la distance maximale entre le plan de classification et les points de données les plus proches. Les points les plus proches de la frontière sont appelés vecteurs de support. La distance orthogonale entre les vecteurs de support et l'hyperplan correspond à la moitié de la marge de la SVM. L'entraînement d'une SVM implique de maximiser la marge en fonction des données des différentes catégories qui se trouvent du côté opposé de l'hyperplan. Les SVM utilisent également des fonctions noyau pour cartographier les données de l'espace d'entrée dans un espace de dimension supérieure (parfois infini), dans lequel l'hyperplan de classification sera choisi.

Ces SVM à marge dure sont rarement utilisées dans la pratique. Un classificateur à marge dure fonctionne uniquement si les données sont linéairement séparables. Lorsqu'un seul échantillon de données se trouve du mauvais côté de l'hyperplan, le classificateur ne peut pas être résolu.

À l'inverse, les classificateurs à marge souple permettent aux échantillons de données d'enfreindre la marge (c'est-à-dire de se situer du mauvais côté de l'hyperplan). Les classificateurs à marge souple tentent d'atteindre une marge maximale tout en limitant les violations de la marge.

La catégorisation de données non étiquetées et l'utilisation dans la prédiction et la reconnaissance des formes sont des exemples d'application des SVM. Lorsque des SVM sont utilisées pour la régression, l'objectif est l'inverse du classificateur SVM. Dans la régression SVM, l'objectif est de faire entrer autant d'instances de données que possible dans la marge, tout en limitant les violations de la marge (les échantillons se trouvant en dehors de la marge).

5.13 Autonomie, hétéronomie et automatisation

Les systèmes d'IA peuvent être comparés en fonction du niveau d'automatisation et du fait qu'ils soient ou non soumis à un contrôle externe. À une extrémité du spectre se trouvent les systèmes autonomes, tandis qu'à l'autre extrémité se trouvent les systèmes entièrement contrôlés par l'homme, ces deux extrémités étant séparées par différents degrés d'hétéronomie. Le [Tableau 1](#) montre la relation entre l'autonomie, l'hétéronomie et l'automatisation, y compris le cas nul d'une absence d'automatisation.

Tableau 1 — Relation entre l'autonomie, l'hétéronomie et l'automatisation

		Niveau d'automatisation	Commentaires
Système automatisé	Autonome	6 - Autonomie	Le système est capable de modifier son domaine d'utilisation ou ses objectifs prévus sans intervention, contrôle ou supervision externe.
	Hétéronome	5 - Automatisation complète	Le système est capable d'effectuer l'intégralité de sa mission sans intervention extérieure.
		4 - Forte automatisation	Le système effectue une partie de sa mission sans intervention extérieure.
		3 - Automatisation conditionnelle	Exécution soutenue et spécifique par un système, un agent externe étant prêt à prendre le contrôle si nécessaire.
		2 - Automatisation partielle	Certaines sous-fonctions du système sont entièrement automatisées tandis que le système reste sous le contrôle d'un agent externe.
		1 - Assistance	Le système aide un opérateur.
		0 - Absence d'automatisation	L'opérateur contrôle entièrement le système.

NOTE Dans la jurisprudence, l'autonomie se rapporte à la capacité d'autogestion. Dans ce sens, également, «autonome» est un terme trompeur lorsqu'il est appliqué aux systèmes d'IA automatisés, car même les systèmes d'IA les plus avancés ne se gèrent pas tout seuls. Les systèmes d'IA fonctionnent au contraire sur la base d'algorithmes et obéissent aux commandes d'opérateurs. Ce sont les raisons pour lesquelles le présent document n'utilise pas le terme populaire d'«autonome» pour décrire l'automatisation^[30].

Les critères pertinents pour la classification d'un système sur ce spectre incluent les critères suivants:

- la présence ou l'absence de supervision externe, soit par un opérateur humain («avec intervention humaine») soit par un autre système automatisé;
- le degré de compréhension contextualisée du système, y compris l'exhaustivité et le caractère opérationnalisable du modèle des états de son environnement du système, et la certitude avec laquelle le système peut raisonner et agir dans son environnement;
- le degré de réactivité ou de réceptivité, y compris si le système peut remarquer des changements dans son environnement, s'il peut réagir aux changements et s'il peut stipuler des changements futurs;
- si son fonctionnement se poursuit jusqu'à ou au-delà de l'accomplissement d'une tâche particulière ou de la survenue d'un événement particulier dans l'environnement (par exemple: pertinent pour la réalisation d'un but, des horodatages ou d'autres mécanismes);
- le degré d'adaptabilité aux changements internes ou externes, aux nécessités ou aux influences;
- la capacité à évaluer ses propres performances ou sa propre adéquation, y compris des évaluations par rapport aux objectifs prédéfinis;
- la capacité à décider et à planifier de façon proactive en ce qui concerne les objectifs, motivations et influences du système.

Au lieu de remplacer du travail humain, dans certains cas la machine viendra le compléter; c'est ce qui est appelé la collaboration homme-machine. Cela peut advenir comme effet secondaire du développement de l'IA, ou un système peut être développé spécifiquement dans le but de créer une collaboration homme-machine. Les systèmes qui visent à compléter les capacités cognitives humaines sont parfois qualifiés d'«augmentation de l'intelligence».

Dans l'ensemble, la présence d'une surveillance redéuable durant l'exploitation peut contribuer à garantir que le système d'IA fonctionne comme prévu et évite des impacts indésirables sur les parties prenantes.

5.14 Internet des objets et systèmes cyber-physiques

5.14.1 Généralités

L'IA est de plus en plus utilisée comme composant dans les systèmes intégrés tels que l'Internet des objets et les systèmes cyber-physiques, soit pour analyser des flux d'informations relatifs au monde physique provenant de capteurs, soit pour faire des prédictions et prendre des décisions au sujet de processus physiques qui sont utilisés pour envoyer des commandes appropriées aux actionneurs afin de contrôler ou d'influencer ces processus physiques.

5.14.2 Internet des objets

L'IdO est une infrastructure composée d'entités, de systèmes et de ressources informationnelles interconnectés, ainsi que de services qui traitent les informations provenant du monde physique et du monde virtuel et y réagissent ([3.1.8](#)). Fondamentalement, un système de l'IdO est un réseau de nœuds doté à la fois de capteurs, qui mesurent les propriétés d'entités physiques puis transmettent des données relatives à ces mesures, et d'actionneurs, qui modifient les propriétés d'entités physiques en réponse à une entrée numérique.

La surveillance médicale et la surveillance de l'état de l'atmosphère sont des exemples de systèmes de l'IdO, dont la sortie est des informations destinées à aider les êtres humains dans leurs activités (par exemple: des avertissements destinés au personnel médical, des prévisions météorologiques pour les humains).

Les systèmes de l'IdO impliquent des systèmes informatiques en réseau qui interagissent avec des entités physiques. Les systèmes de l'IdO s'appuient sur des dispositifs IdO numériques, sous la forme de capteurs et d'actionneurs, qui interagissent avec des entités physiques. Un capteur mesure une ou plusieurs propriétés d'une ou plusieurs entités physiques et produit des données qui peuvent être transmises sur un réseau. Un actionneur modifie une ou plusieurs propriétés d'une entité physique en réponse à une entrée valide, reçue sur un réseau. Les capteurs et les actionneurs peuvent se présenter sous de nombreuses formes, telles que des thermomètres, des accéléromètres, des caméras vidéo, des microphones, des relais, des dispositifs de chauffage, des robots ou des équipements industriels pour la fabrication ou le contrôle de processus. Voir l'ISO/IEC 30141 pour plus d'informations.

L'IA peut jouer un rôle important dans le contexte des systèmes de l'IdO. Cela comprend l'analyse des données entrantes et la prise de décision qui peut aider à atteindre les objectifs du système, tels que le contrôle des entités physiques et des processus physiques, en fournissant des informations contextuelles, en temps réel et prédictives.

5.14.3 Systèmes cyber-physiques

Dans les CPS, qui sont des systèmes similaires à l'IdO, une logique de commande est appliquée à l'entrée provenant des capteurs afin de diriger les activités des actionneurs et par conséquent influencer les processus ayant lieu dans le monde physique.

Un robot est un exemple de système CPS, où l'entrée provenant des capteurs est directement utilisée pour commander les activités du robot et effectuer des actions dans le monde physique.

La robotique englobe toutes les activités relatives à la conception, à l'assemblage, à la production, au contrôle et à l'utilisation de robots pour différents types d'applications. Un robot est constitué de composants électroniques, mécaniques, micrologiciels et logiciels qui interagissent étroitement entre eux afin d'atteindre les objectifs définis pour une application spécifique. Les robots sont généralement équipés de capteurs afin d'évaluer leur situation actuelle, de processeurs afin d'offrir un contrôle au travers de l'analyse et de la planification d'actions, et d'actionneurs permettant de réaliser les actions. Les robots industriels installés dans les cellules de fabrication sont programmés pour répéter de façon

précise les mêmes trajectoires et actions, encore et encore, sans faire d'écart. Les robots de service ou les robots collaboratifs ont besoin de s'adapter à des situations changeantes et à des environnements dynamiques. Programmer cette flexibilité est quasiment impossible en raison de toute la variabilité que cela implique.

Les composants du système d'IA peuvent servir dans le cadre du logiciel de commande et du processus de planification au travers du paradigme «déetecter, planifier, agir», ce qui permet aux robots de s'adapter lorsqu'apparaissent des obstacles ou que des objets cibles se sont déplacés. Associer la robotique et des composants de système d'IA permet une interaction physique automatisée avec les objets, l'environnement et les personnes.

5.15 Crédibilité

5.15.1 Généralités

La crédibilité des systèmes d'IA se rapporte aux caractéristiques qui aident les parties prenantes pertinentes à comprendre si le système d'IA répond ou non à leurs attentes. Ces caractéristiques peuvent aider les parties prenantes à vérifier que:

- les systèmes d'IA ont été correctement conçus et validés conformément aux règles et normes représentant l'état de l'art. Cela implique la garantie de la qualité et de la robustesse;
- les systèmes d'IA sont construits pour le bénéfice des parties prenantes pertinentes, qui ont des objectifs harmonisés. Cela implique une sensibilisation au fonctionnement des algorithmes d'IA et une compréhension du fonctionnement global par les parties prenantes. Cela implique également une qualification ou une assurance de la certification du développement et de l'exploitation de l'IA conformément aux exigences légales et aux normes sectorielles lorsqu'elles existent;
- les systèmes d'IA sont fournis avec une identification adéquate des parties responsables et redevables;
- les systèmes d'IA sont développés et exploités en tenant compte des préoccupations régionales appropriées.

Pour plus d'informations, voir l'ISO/IEC TR 24028.

5.15.2 Robustesse de l'IA

Pour les systèmes d'IA, la robustesse peut décrire leur capacité à maintenir leur niveau de performance, tel que prévu par leurs développeurs, dans toutes les situations. Un exemple de robustesse est la capacité d'un système à fonctionner dans les limites acceptables en dépit de conditions environnementales externes ou difficiles. La robustesse peut englober d'autres attributs tels que la résilience et la fiabilité. Le bon fonctionnement d'un système d'IA est lié à, ou peut permettre, la sécurité de ses parties prenantes dans un environnement ou un contexte donnés (voir l'ISO/IEC TR 24028).

Par exemple, un système d'IA robuste basé sur le ML peut être en capacité de généraliser à des entrées bruitées, telles que l'absence de surajustement. Pour parvenir à la robustesse, une option consiste à entraîner le(s) modèle(s) à l'aide de grands ensembles de données d'entraînement incluant des données d'entraînement bruitées (voir l'ISO/IEC TR 24028).

Les propriétés de robustesse démontrent la capacité (ou l'incapacité) du système à afficher des performances comparables sur des données atypiques par opposition aux données attendues dans les opérations types, ou sur des entrées différentes de celles sur lesquelles il a été entraîné (voir l'ISO/IEC TR 24029-1).

Lors du traitement de données d'entrée, un système d'IA est censé générer des prédictions (ses sorties) dans une plage acceptable, cohérente et efficace. Même si ces sorties ne sont pas idéales, un système peut toujours être considéré comme robuste. Un système d'IA dont les sorties ne se situent pas dans cette plage acceptable, cohérente ou efficace lors du traitement de données d'entrée ne peut pas être considéré comme robuste.

La robustesse peut être envisagée différemment pour différents types de systèmes d'IA, tels que:

- la robustesse d'un modèle de régression est la capacité à avoir des mesures acceptables d'amplitude de réponse sur toute entrée valide;
- la robustesse d'un modèle de classification est la capacité à éviter d'insérer de nouvelles erreurs de classification lors du passage d'entrées types à des entrées se situant dans une certaine plage par rapport à celles-ci.

5.15.3 Fiabilité de l'IA

La fiabilité est la capacité d'un système ou d'une entité de ce système à exécuter ses fonctions requises dans les conditions déclarées pendant une période spécifique (voir l'ISO/IEC 27040).

La fiabilité d'un système d'IA fait référence à la capacité qui lui permet de fournir continuellement la prédiction ([3.1.27](#)), la recommandation et la décision correctes requises lors de son étape d'exploitation ([6.2.6](#)).

La fiabilité peut être affectée au moins par la robustesse, la généralisabilité, la cohérence et la résilience d'un système d'IA. Toutes les entrées et tous les paramètres de l'environnement qui satisfont aux critères énoncés sont censés être traités correctement durant son fonctionnement. Certaines des entrées peuvent être différentes de celles utilisées lors de son étape de développement, mais elles peuvent se produire lorsque le système est utilisé. La sauvegarde d'un composant ou d'un système d'IA améliore également la fiabilité, ce qui permettrait d'obtenir des mises en œuvre de la logique métier qui se comportent de la même manière que l'original. Elle fonctionne lorsque le système d'IA connaît une défaillance.

La fiabilité peut soutenir la sécurité fonctionnelle d'un système d'IA, dans le sens où des protections et des opérations automatiques en cas de défaillance définie sont requises (par les parties prenantes) pour le système ou pour une partie de celui-ci.

5.15.4 Résilience de l'IA

La résilience est la capacité du système à retrouver son état opérationnel rapidement après un incident. La tolérance aux pannes est la capacité du système à continuer à fonctionner en cas de survenue de perturbations, de pannes et de défaillances au sein du système, éventuellement avec des capacités dégradées.

Avec les systèmes d'IA, comme avec d'autres types de systèmes logiciels, les défaillances matérielles peuvent avoir une incidence sur l'exécution correcte de l'algorithme.

La fiabilité est liée à la résilience, mais les niveaux de service attendus et les attentes sont différents, les attentes en matière de résilience étant éventuellement moins élevées telles que définies par les parties prenantes. Les systèmes résilients peuvent offrir un niveau de fonctionnement dégradé face à certains types de défaillance, ce qui peut être acceptable pour les parties prenantes. Il convient que les systèmes résilients incluent également des approches en matière de récupération en cas de besoin.

5.15.5 Contrôlabilité de l'IA

La contrôlabilité est la propriété d'un système d'IA grâce à laquelle un agent externe peut intervenir dans son fonctionnement. La contrôlabilité peut être obtenue en fournissant des mécanismes fiables via lesquels un agent peut prendre le contrôle du système d'IA.

Un aspect clé de la contrôlabilité consiste à déterminer quel(s) agent(s) peut (peuvent) contrôler quels composants du système d'IA (par exemple: le fournisseur de service ou les vendeurs de produits, le fournisseur de l'IA constitutive, l'utilisateur ou une entité dotée de pouvoirs réglementaires).

De plus amples informations sur la contrôlabilité sont disponibles dans l'ISO/IEC TR 24028:2020, 10.4.

5.15.6 Explicabilité de l'IA

L'explicabilité est la propriété d'un système d'IA qui signifie que les facteurs importants qui influencent une décision peuvent être exprimés d'une manière que les humains peuvent comprendre. L'explicabilité peut être particulièrement importante lorsque les décisions prises par un système d'IA ont une incidence sur un ou plusieurs êtres humains. Les êtres humains sont susceptibles de se méfier d'une décision s'ils ne peuvent pas comprendre comment elle a été prise, en particulier lorsque la décision leur est d'une façon ou d'une autre défavorable sur un plan personnel (par exemple: refus d'une demande de prêt).

L'explicabilité peut également être un moyen utile de valider le système d'IA, même lorsque les décisions n'ont pas d'incidence directe sur les êtres humains. Par exemple, si un système d'IA analyse l'image d'une scène pour identifier des entités présentes dans cette scène, il peut être utile de voir une explication des raisons d'une décision prise au sujet du contenu de la scène, comme moyen de vérifier que ce qui est identifié est bien ce qui est revendiqué. Il existe des contre-exemples dans l'histoire des systèmes d'IA où de telles explications n'étaient pas disponibles et il a été établi que le système d'IA identifiait certaines entités dans une scène sur la base de corrélations erronées qui existaient dans les données d'entraînement.

L'explicabilité peut être plus facile pour certains types de systèmes d'IA que pour d'autres. Les réseaux neuronaux d'apprentissage profond peuvent être problématiques, car la complexité du système peut rendre difficile la fourniture d'une explication significative sur la façon dont le système arrive à une décision.

Les algorithmes basés sur des règles, tels que les méthodes symboliques ou les arbres de décision, sont souvent considérés comme très explicables, car ces règles fournissent directement une explication. Cependant, l'explication peut être moins compréhensible lorsque la taille et la complexité de ces modèles augmentent.

5.15.7 Prévisibilité de l'IA

La prévisibilité est la propriété d'un système d'IA qui permet aux parties prenantes de formuler des hypothèses fiables au sujet de la sortie. La prévisibilité joue un rôle important dans l'acceptabilité des systèmes d'IA et est souvent mentionnée dans les débats consacrés à l'éthique en ce qui concerne les systèmes d'IA. La confiance dans la technologie repose souvent sur la prévisibilité: un système inspire confiance si les utilisateurs peuvent déduire comment le système se comportera dans une situation particulière, même si les utilisateurs ne peuvent pas expliquer les facteurs qui président au comportement du système. Au contraire, les utilisateurs peuvent arrêter de faire confiance à un système si celui-ci commence à fonctionner de manière surprenante dans des situations où la bonne réponse semble évidente.

Toutefois, la notion naïve de prévisibilité basée sur l'idée qu'il convient qu'un être humain soit capable de prédire le comportement d'un système d'IA pose plusieurs problèmes:

- une définition directement basée sur la compréhension humaine est intrinsèquement subjective. Il convient qu'une définition de la prévisibilité utilise des critères objectifs et quantifiables;
- il convient qu'il soit possible de faire confiance à un système d'IA même si un humain ne peut pas prédire son comportement précis dans toutes les situations. Une garantie statistique de l'adéquation de son comportement peut être plus utile. La justification de cette affirmation est que de nombreuses approches d'apprentissage machine produisent des résultats nécessairement imprévisibles.

La prévisibilité est associée à l'exactitude. Les méthodes visant à améliorer l'exactitude peuvent réduire la probabilité que les systèmes d'IA produisent des résultats imprévisibles.

5.15.8 Transparence de l'IA

La transparence des systèmes d'IA soutient les objectifs centrés sur l'humain pour le système et constitue un sujet de recherche et de discussion permanent. Offrir une transparence au sujet d'un système d'IA peut impliquer de communiquer aux parties prenantes des informations appropriées sur le système (par exemple: objectifs, limites connues, définitions, choix de conception, hypothèses,

caractéristiques, modèles, algorithmes, méthodes d'entraînement et processus d'assurance qualité). De plus, la transparence d'un système d'IA peut impliquer d'informer les parties prenantes au sujet des détails des données utilisées (par exemple: quoi, où, quand, pourquoi les données sont collectées et comment elles sont utilisées) pour produire le système et la protection des données personnelles, ainsi que sur l'objectif du système et la manière dont il a été construit et déployé. La transparence peut également impliquer d'informer les parties prenantes au sujet du traitement et du niveau d'automatisation utilisé pour prendre les décisions associées.

NOTE Cette divulgation de certaines informations dans une quête de transparence peut aller à l'encontre des exigences en matière de sécurité, de protection de la vie privée ou de confidentialité.

5.15.9 Biais et équité de l'IA

De manière générale, la signification du terme «biais» dépend du contexte.

Dans le domaine de l'IA, le terme «biais» fait référence à l'idée que des cas différents imposent un traitement différent. Dans ce sens, le biais est ce qui permet aux systèmes d'apprentissage machine de juger qu'une situation est différente d'une autre et de se comporter différemment en conséquence. En tant que tel, le biais est essentiel pour le processus d'apprentissage machine et pour l'adaptation du comportement à la situation particulière considérée.

Dans le contexte social, toutefois, le terme «biais» se rapporte souvent à la notion que certaines différences de traitement sont injustes. Pour éviter toute confusion, dans le contexte de l'IA, plutôt que biais, le terme d'iniquité est utilisé pour désigner une différence de traitement injustifiée qui profite prioritairement à certains groupes plutôt qu'à d'autres. Le comportement injuste d'un système d'IA peut conduire au non-respect de faits, de croyances et de normes établis, ce qui mène au favoritisme ou à la discrimination.

Bien que certains biais soient essentiels au bon fonctionnement d'un système d'IA, un biais indésirable peut être introduit dans un système d'IA de manière involontaire et peut conduire à des résultats injustes du système. Les sources de biais indésirable dans les systèmes d'IA sont interdépendantes et incluent le biais cognitif humain, le biais dans les données et le biais introduit par les décisions techniques. Le biais dans les données d'entraînement représente la principale source de biais dans les systèmes d'IA. Les biais cognitifs humains peuvent avoir une incidence sur les décisions relatives à la collecte et au traitement des données, à la conception du système, à l'entraînement des modèles et à d'autres décisions en matière de développement.

Minimiser le biais indésirable dans les systèmes d'IA est un objectif difficile, mais sa détection et son traitement sont possibles. [ISO/IEC TR 24027:2021].

5.16 Vérification et validation de l'IA

La vérification est la confirmation qu'un système a été construit correctement et qu'il satisfait aux exigences spécifiées. La validation est la confirmation, par la fourniture de preuves objectives, que les exigences relatives à une utilisation prévue ou une application prévue ont été satisfaites. Les considérations relatives à la vérification et à la validation incluent les suivantes:

- certains systèmes sont totalement vérifiables (par exemple: tous les composants du système peuvent être vérifiés individuellement, tout comme le système complet);
- certains systèmes sont partiellement vérifiables et partiellement validables (par exemple: au moins un composant du système peut être vérifié individuellement et les composants restants du système peuvent être validés, tout comme le système complet);
- certains systèmes sont invérifiables mais validables (par exemple: aucun composant du système ne peut être vérifié, mais tous les composants du système peuvent être validés, tout comme le système complet);

- certains systèmes sont invérifiables et partiellement validables (par exemple: aucun composant du système ne peut être vérifié, mais au moins un composant du système peut être validé individuellement);
- certains systèmes sont invérifiables et non validables (par exemple: aucun composant du système ne peut être vérifié ou validé).

5.17 Questions juridictionnelles

Les systèmes d'IA peuvent être déployés et exploités dans des juridictions autres que celles dans lesquelles le système a été conçu ou fabriqué. Il convient que les développeurs et les fabricants de systèmes d'IA aient conscience du fait que les exigences légales applicables peuvent varier d'une juridiction à l'autre.

Par exemple, une voiture fabriquée dans une juridiction peut être tenue de se conformer à des exigences légales différentes pour être autorisée à entrer dans une autre juridiction.

En outre, les systèmes d'IA nécessitent généralement la collecte, le traitement et l'utilisation de données au cours des étapes de développement et d'exploitation du système d'IA et leur élimination au cours de la phase de mise hors service. Il convient que les développeurs, fabricants et utilisateurs de systèmes d'IA aient conscience du fait que les exigences légales en matière de collecte, d'utilisation et de suppression des données peuvent également varier d'une juridiction à l'autre.

Pour atténuer l'impact de la variation des exigences légales, les développeurs et les fabricants de systèmes d'IA peuvent utiliser une ou plusieurs des mesures d'atténuation suivantes:

- noter les exigences légales qui peuvent s'appliquer au système d'IA pendant la phase de préparation. Il convient que cela inclue également les exigences légales relatives à la collecte, à l'utilisation et à la suppression des données;
- élaborer un plan pour se conformer aux exigences légales applicables dans la (les) juridiction(s) dans laquelle (lesquelles) le système d'IA est destiné à être déployé et exploité;
- élaborer un plan de surveillance de la conformité aux exigences légales durant les étapes de développement, de déploiement, d'exploitation et de mise hors service du système d'IA;
- élaborer un plan de surveillance et de réaction à toute modification des exigences légales;
- adopter des approches flexibles en matière de conception, de déploiement et d'exploitation.

5.18 Impact sociétal

Les systèmes d'IA présentent un éventail de risques, déterminé par la gravité de l'impact potentiel d'une défaillance ou d'un comportement inattendu. Les aspects pertinents pour l'évaluation du niveau de risque sont les suivants:

- le type d'espace de travail dans lequel le système opère (par exemple: recommandations par opposition à l'action directe dans un environnement);
- la présence ou l'absence de supervision externe;
- le type de supervision externe (automatisée ou manuelle);
- la pertinence éthique de la tâche ou du domaine;
- le niveau de transparence des décisions ou des étapes de traitement;
- le degré d'automatisation du système.

Par exemple, un système qui ne formule que des recommandations et ne peut pas agir seul, dans un domaine qui n'a pas de pertinence éthique, peut être considéré comme à faible risque. À l'inverse, un système d'IA peut être considéré comme présentant un risque négatif élevé si ses actions ont un impact

direct sur des vies humaines, s'il fonctionne sans supervision externe et si sa prise de décisions est opaque.

NOTE Pour des domaines d'application spécifiques, des politiques, des normes et des exigences légales supplémentaires peuvent s'appliquer et aller au-delà de l'analyse d'impact décrite dans le présent paragraphe.

5.19 Rôles des parties prenantes de l'IA

5.19.1 Généralités

Tel qu'illustré à la [Figure 2](#), l'IA peut impliquer plusieurs rôles et sous-rôles de parties prenantes. Ces rôles et sous-rôles sont décrits aux [Paragraphes 5.19.2 à 5.19.7](#).

NOTE Un organisme ou une entité peut assumer plus d'un rôle ou sous-rôle.

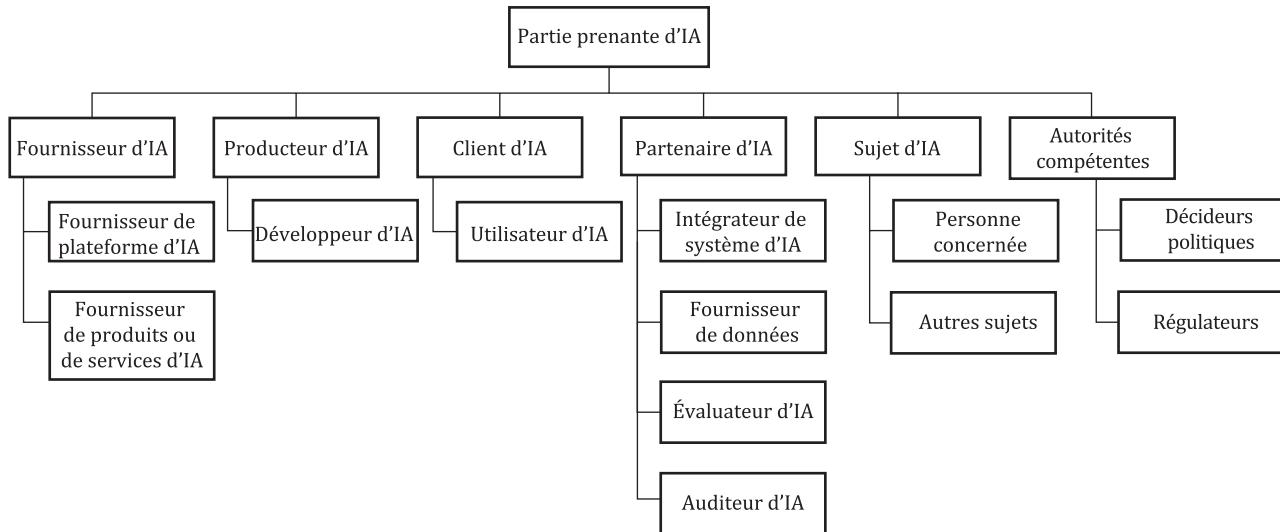


Figure 2 — Rôles des parties prenantes de l'IA et sous-rôles associés

5.19.2 Fournisseur d'IA

5.19.2.1 Généralités

Un fournisseur d'IA est un organisme ou une entité qui fournit des produits ou des services qui utilisent un ou plusieurs systèmes d'IA. Les fournisseurs d'IA englobent les fournisseurs de plateforme d'IA et les fournisseurs de produit ou de service d'IA.

5.19.2.2 Fournisseur de plateforme d'IA

Un fournisseur de plateforme d'IA est un organisme ou une entité qui fournit des services qui permettent à d'autres parties prenantes de produire des services ou des produits d'IA.

5.19.2.3 Fournisseur de produit ou de service d'IA

Un fournisseur de produit ou de service d'IA est un organisme ou une entité qui fournit des produits ou des services d'IA qui sont directement utilisables par un client ou un utilisateur d'IA, ou qui sont à intégrer dans un système qui utilise l'IA conjointement avec des composants non IA.

5.19.3 Producteur d'IA

5.19.3.1 Généralités

Un producteur d'IA est un organisme ou une entité qui conçoit, qui développe, qui soumet à l'essai et qui déploie des produits ou des services qui utilisent un ou plusieurs systèmes d'IA.

5.19.3.2 Développeur d'IA

Un développeur d'IA est un organisme ou une entité qui s'attache au développement de produits et de services d'IA. Les exemples de développeurs d'IA incluent, sans que cela s'y limite:

- concepteur de modèle: l'entité qui reçoit des données et une spécification de problème et crée un modèle d'IA;
- implémenteur de modèle: l'entité qui reçoit un modèle d'IA et qui spécifie le calcul à exécuter (la mise en œuvre à utiliser et sur quelles ressources de calcul, par exemple: CPU, GPU, ASIC, FPGA);
- vérificateur de calcul: l'entité qui vérifie qu'un calcul est exécuté tel que prévu;
- vérificateur de modèle: l'entité qui vérifie que le modèle d'IA fonctionne comme prévu.

5.19.4 Client d'IA

5.19.4.1 Généralités

Un client d'IA est un organisme ou une entité qui utilise un produit ou un service d'IA, soit directement, soit en le fournissant à des utilisateurs d'IA.

5.19.4.2 Utilisateurs d'IA

Un utilisateur d'IA est un organisme ou une entité qui utilise des produits ou des services d'IA.

5.19.5 Partenaire d'IA

5.19.5.1 Généralités

Un partenaire d'IA est un organisme ou une entité qui fournit des services dans le contexte de l'IA. Les partenaires d'IA peuvent réaliser le développement technique de produits ou de services d'IA, procéder à des essais et à la validation de produits et de services d'IA, réaliser des audits de l'utilisation de l'IA, évaluer des produits ou des services d'IA et effectuer d'autres tâches. Des exemples de types de partenaires d'IA sont abordés dans les paragraphes suivants.

5.19.5.2 Intégrateur de système d'IA

Un intégrateur de système d'IA est un organisme ou une entité qui s'occupe de l'intégration de composants d'IA dans des systèmes plus étendus, qui éventuellement comprennent également des composants non IA.

5.19.5.3 Fournisseur de données

Un fournisseur de données est un organisme ou une entité qui s'occupe de fournir les données utilisées par des produits ou des services d'IA.

5.19.5.4 Auditeur d'IA

Un auditeur d'IA est un organisme ou une entité qui s'occupe de réaliser l'audit des organismes qui produisent, fournissent ou utilisent des systèmes d'IA, afin d'évaluer la conformité aux normes, aux politiques ou aux exigences légales.

5.19.5.5 Évaluateur d'IA

Un évaluateur d'IA est un organisme ou une entité qui évalue les performances d'un ou de plusieurs systèmes d'IA.

5.19.6 Sujet d'IA

5.19.6.1 Généralités

Un sujet d'IA est un organisme ou une entité qui est affecté(e) par un système, un service ou un produit d'IA.

5.19.6.2 Personne concernée

Une personne concernée est un organisme ou une entité qui est affecté(e) par les systèmes d'IA avec les aspects suivants:

- personne concernée impliquée dans l'entraînement: lorsque des données relatives à un organisme ou à un être humain sont utilisées pour l'entraînement d'un système d'IA, il peut y avoir des implications pour la sécurité et la protection de la vie privée, sur ce second aspect particulièrement quand le sujet est un être humain.

5.19.6.3 Autres personnes concernées

Les autres organismes ou entités affectés par un système, un service ou un produit d'IA peuvent être, par exemple, une personne ou une communauté. Par exemple, les consommateurs qui interagissent avec un réseau social qui fournit des recommandations basées sur l'IA, les conducteurs de véhicules équipés d'une automatisation basée sur l'IA.

5.19.7 Autorités compétentes

5.19.7.1 Généralités

Les autorités compétentes sont des organismes ou des entités qui peuvent avoir un impact sur un système, un service ou un produit d'IA.

5.19.7.2 Décideurs politiques

Il s'agit d'organismes et d'entités qui sont habilités à définir des politiques dans un domaine international, régional, national ou industriel qui peuvent avoir un impact sur un système, un service ou un produit d'IA.

5.19.7.3 Régulateurs

Il s'agit d'organismes et d'entités qui sont habilités à définir, mettre en œuvre et faire respecter les exigences légales telles que prévues dans les politiques définies par les décideurs politiques (5.179.2).

6 Cycle de vie de système d'IA

6.1 Modèle de cycle de vie de système d'IA

Le modèle de cycle de vie de système d'IA décrit l'évolution d'un système d'IA depuis son initialisation jusqu'à sa mise hors service. Le présent document ne prescrit pas de modèle de cycle de vie spécifique, mais souligne certains processus spécifiques aux systèmes d'IA qui peuvent survenir au cours du cycle de vie du système. Des processus et des jalons spécifiques peuvent survenir au cours d'une ou de plusieurs étapes du cycle de vie et certaines étapes du cycle de vie peuvent être répétées durant l'existence du système. Par exemple, il peut être décidé de répéter plusieurs fois les étapes de «conception et développement» et de «déploiement» afin de développer et mettre en œuvre des correctifs de bogues et des mises à jour du système.

Un modèle de cycle de vie de système aide les parties prenantes à construire des systèmes d'IA de manière plus efficace et efficiente. Les normes internationales sont utiles pour élaborer le modèle de cycle de vie, notamment l'ISO/IEC 15288 pour les systèmes dans leur ensemble, l'ISO/IEC 12207 pour les logiciels et l'ISO/IEC 15289 pour la documentation des systèmes. Ces normes internationales décrivent les processus du cycle de vie pour des systèmes généraux, non spécifiques aux systèmes d'IA. La [Figure 3](#) fournit un exemple des étapes et des processus de haut niveau qui peuvent être appliqués au cycle de vie d'un système d'IA. Les étapes et les processus peuvent être exécutés de manière itérative, ce qui est souvent nécessaire pour le développement et l'exploitation d'un système d'IA. Il existe diverses considérations qu'il convient de prendre en compte dans l'ensemble du cycle de vie, du début à la fin. Ces considérations incluent, par exemple:

- les implications en matière de gouvernance découlant du développement ou de l'utilisation de systèmes d'IA;
- les implications en matière de protection de la vie privée et de sécurité dues à l'utilisation de grandes quantités de données, dont certaines peuvent être de nature sensible;
- les menaces pour la sécurité qui découlent du développement de systèmes dépendant des données;
- les aspects liés à la transparence et à l'explicabilité, notamment la provenance des données et la capacité à fournir une explication sur la manière dont les sorties d'un système d'IA sont déterminées.

La [Figure 3](#) montre un exemple de processus de haut niveau et d'étapes du modèle de cycle de vie d'un système d'IA. L'[Annexe A](#) montre comment les correspondances entre ce modèle de cycle de vie de système d'IA et la définition du cycle de vie d'un système d'IA de l'OCDE.

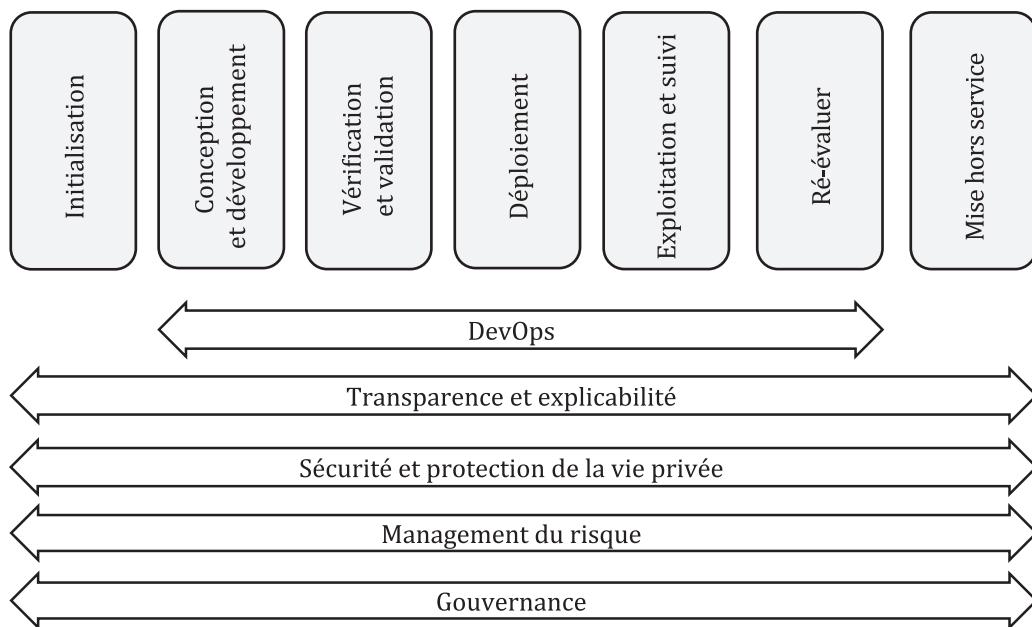


Figure 3 — Exemple de processus de haut niveau et d'étapes de modèle de cycle de vie d'un système d'IA

Les systèmes d'IA diffèrent des autres types de systèmes, ce qui peut avoir une incidence sur les processus du modèle de cycle de vie. Par exemple:

- la plupart des systèmes logiciels sont programmés pour se comporter selon des manières précisément définies qui sont dictées par leurs exigences et leurs spécifications. Les systèmes d'IA basés sur l'apprentissage machine utilisent des méthodes d'entraînement et d'optimisation basées sur les données pour traiter des entrées très variables;
- les applications logicielles traditionnelles sont généralement prévisibles, ce qui est moins souvent le cas des systèmes d'IA;
- les applications logicielles traditionnelles sont également généralement vérifiables, tandis que l'évaluation des performances des systèmes d'IA nécessite souvent des approches statistiques et leur vérification peut être difficile;
- les systèmes d'IA ont généralement besoin de multiples itérations d'amélioration pour atteindre des niveaux de performance acceptables.

La gestion des données (qui englobe les processus et les outils pour l'acquisition de données, l'annotation des données, la préparation des données, la vérification de la qualité des données, l'échantillonnage des données et l'augmentation des données) est un aspect clé des systèmes d'IA.

Les processus de développement et d'essai diffèrent également pour les systèmes d'IA, étant donné que ces processus sont également basés sur des données. Cela devient plus difficile pour les systèmes d'IA qui utilisent l'apprentissage continu (également appelé apprentissage permanent ou apprentissage tout au long de la vie), où le système apprend durant l'étape d'exploitation et où des essais continus sont requis.

Le processus de gestion des versions des systèmes d'IA est différent de celui des logiciels traditionnels. Alors que les applications logicielles traditionnelles traitent des fonctions de gestion de version de code et de différence de code, les versions des systèmes d'IA incluent les différences de code et de modèle, ainsi que les différences de données d'entraînement si l'apprentissage machine est utilisé.

Certains des processus du cycle de vie de l'IA qui diffèrent de ceux du cycle de vie des logiciels traditionnels sont abordés en [6.2](#).

La [Figure 4](#) montre un exemple de modèle de cycle de vie pour un système d'IA; des modèles de cycle de vie différents sont possibles en fonction des différentes techniques de développement. La [Figure 4](#) montre une série d'étapes du cycle de vie et indique certains des processus de chaque étape qui sont importants pour les systèmes d'IA et qui nécessitent une attention particulière, au-delà de l'attention requise pour le développement de systèmes non IA types.

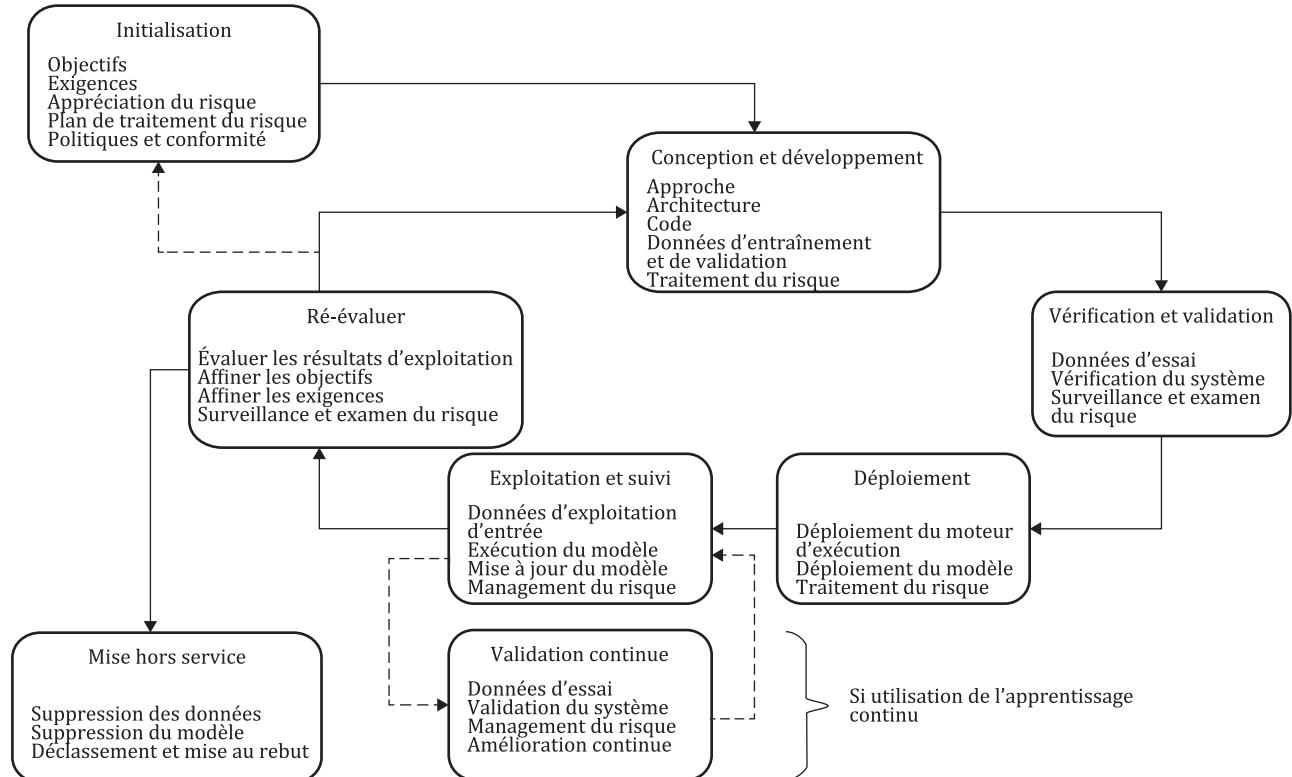


Figure 4 — Exemple de modèle de cycle de vie d'un système d'IA avec des processus spécifiques au système d'IA

Tel que représenté à la [Figure 4](#), le développement et l'exploitation des systèmes d'IA ont tendance à être de nature plus itérative que pour les systèmes non IA. Les systèmes d'IA ont tendance à être moins prévisibles et il faut généralement une certaine expérience opérationnelle et un ajustement du système d'IA pour qu'il atteigne ses objectifs.

6.2 Processus et étapes du cycle de vie d'un système d'IA

6.2.1 Généralités

Les processus décrits à chaque étape sont des exemples représentatifs, car les processus spécifiques dépendront du système d'IA. Les processus peuvent être exécutés dans des ordres différents et, dans certains cas, en parallèle.

Ces processus ne sont pas nécessairement eux-mêmes spécifiques à l'IA, mais les enjeux associés à l'IA leur confèrent une importance particulière dans ce contexte.

6.2.2 Initialisation

L'initialisation intervient lorsqu'une ou plusieurs parties prenantes décident de transformer une idée en un système concret. L'étape d'initialisation peut impliquer plusieurs processus et décisions qui conduisent à la décision de passer à l'étape de conception et de développement. L'étape d'initialisation peut être revue durant le cycle de vie si de nouvelles informations sont découvertes à des étapes ultérieures. Par exemple, il est possible de découvrir que le système n'est pas techniquement ou

financièrement réalisable. Les processus qui peuvent se dérouler durant l'étape d'initialisation incluent par exemple:

Objectifs: il convient que les parties prenantes déterminent pourquoi il est nécessaire d'élaborer un système d'IA. Quel problème le système permet-il de résoudre? À quel besoin du client ou à quelle opportunité commerciale le système répond-il? Quelles sont les mesures de la réussite?

Exigences: il convient que les parties prenantes préparent pour le système d'IA un ensemble d'exigences qui couvre tout le cycle de vie du système d'IA. Le fait de ne pas prendre en compte les exigences relatives au déploiement, à l'exploitation et à la mise hors service peut entraîner des problèmes à l'avenir. Une approche associant toutes les parties prenantes et incluant diverses expertises thématiques peut contribuer à identifier les risques potentiels et les conséquences imprévues du système. Il convient que les parties prenantes s'assurent que les exigences du système d'IA répondent à ses objectifs. Il convient que les exigences tiennent compte du fait que de nombreux systèmes d'IA ne sont pas prévisibles et de l'impact que cela peut avoir sur la réalisation des objectifs. Il convient que les parties prenantes prennent en compte les facteurs réglementaires et s'assurent que le développement et l'exploitation du système d'IA sont conformes aux politiques obligatoires correspondantes.

Management du risque: il convient que les organismes apprécient les risques liés à l'IA durant tout le cycle de vie d'un système d'IA. Il convient que le résultat de cette activité soit un plan de traitement du risque. Le management du risque, y compris l'identification, l'appréciation et le traitement du risque liés à l'IA sont décrits dans l'ISO/IEC 23894.

Il convient que les organismes identifient les dommages et les avantages potentiels liés au système d'IA, y compris en se concertant avec des utilisateurs représentatifs. Ce processus peut produire un ensemble de valeurs qui peuvent guider le développement de certaines parties du système, y compris les caractéristiques, l'interface utilisateur, la documentation et les utilisations. Il convient que les organismes étudient et affinent ces valeurs jusqu'à ce qu'elles puissent faire partie des exigences du système. Les cadres juridiques, liés aux droits de l'homme, à la responsabilité sociale et à l'environnement peuvent contribuer à affiner et à décrire les valeurs.

Outre les risques habituels pris en compte pour un système, tels que la sécurité et la protection de la vie privée, il convient que le plan couvre également les risques liés aux valeurs identifiées.

Transparence et redevabilité: il convient que les parties prenantes veillent à ce que, tout au long du cycle de vie, les considérations telles que la provenance des données, la validité des sources de données, les efforts d'atténuation des risques, les processus et les décisions mis en œuvre soient consignés afin de permettre une compréhension exhaustive de la manière dont les résultats du système d'IA sont déduits, ainsi qu'à des fins de redevabilité.

Coût et financement: il convient que les parties prenantes prévoient les coûts du système d'IA tout au long du cycle de vie et s'assurent qu'un financement est disponible.

Ressources: il convient que les parties prenantes déterminent quelles ressources sont nécessaires pour mettre en œuvre et compléter chaque étape du cycle de vie et il est nécessaire qu'elles s'assurent que les ressources seront disponibles en cas de besoin. Il convient de prendre en compte les données qui peuvent être nécessaires au développement ou à l'évaluation d'un système d'IA. Pour les systèmes d'IA basés sur le ML, il convient d'accorder une attention particulière aux données d'entraînement, de validation et d'essai.

Faisabilité: l'étape d'initialisation donne lieu à une décision établissant si le système d'IA est réalisable. Une démonstration de faisabilité peut être réalisée afin de déterminer si le système satisfait aux exigences et aux objectifs. Les exigences et les objectifs peuvent inclure par exemple:

- répond au problème défini;
- répond à une opportunité commerciale ou remplit une mission;
- satisfait aux capacités et attributs spécifiés.

Si le système d'IA est jugé réalisable, les parties prenantes peuvent décider de passer à l'étape de développement.

6.2.3 Conception et développement

L'étape de conception et de développement crée le système d'IA et se termine avec un système d'IA qui est prêt pour la vérification et la validation. Au cours de cette étape, et notamment avant la conclusion, il convient que les parties prenantes s'assurent que le système d'IA satisfait aux objectifs, aux exigences et aux autres cibles initiaux identifiés lors de l'étape d'initialisation. Les processus qui peuvent se dérouler durant l'étape de conception et de développement incluent par exemple:

Approche: il convient que les parties prenantes déterminent une approche globale pour concevoir le système d'IA, le soumettre à l'essai et le rendre prêt à être réceptionné et déployé. L'étape d'approche peut inclure une réflexion pour déterminer si à la fois du matériel et des logiciels seront nécessaires et identifier où obtenir les composants (par exemple: développer à partir de zéro, acheter du matériel disponible dans le commerce, utiliser des logiciels libres).

Architecture: il convient que les parties prenantes déterminent et documentent l'architecture globale du système d'IA. Les processus liés à l'architecture et l'approche sont liés et il peut être nécessaire d'itérer entre les deux.

Code: le code logiciel du système d'IA est développé ou acquis.

Données d'entraînement: les systèmes d'IA incorporent les connaissances acquises. Le traitement des données d'entraînement est un élément fondamental du développement des systèmes d'IA basés sur l'apprentissage machine (voir [5.10](#)).

Traitement du risque: il convient que les organismes mettent en œuvre les processus et les contrôles décrits dans le plan de traitement du risque (voir l'[ISO/IEC 23894](#)).

6.2.4 Vérification et validation

La vérification et la validation vérifient que le système d'IA issu de l'étape de conception et de développement fonctionne conformément aux exigences et atteint les objectifs.

Les processus qui peuvent faire partie de la vérification et de la validation incluent par exemple:

Vérification: le logiciel est soumis à l'essai afin de vérifier sa fonctionnalité et la présence de bogues, comme tout matériel. Des essais d'intégration des systèmes peuvent également être effectués. Un essai de performance peut être effectué, et vérifier si le temps de réponse, le délai ou toute autre caractéristique de performance pertinente du système d'IA satisfait aux exigences spécifiques.

Un aspect important des systèmes d'IA est la nécessité de vérifier que les capacités d'IA fonctionnent tel que prévu. Cela nécessite l'acquisition, la préparation et l'utilisation de données d'essai. Il est nécessaire que les données d'essai soient distinctes de toutes les autres données utilisées lors de la conception et du développement et qu'elles soient également représentatives des données d'entrée que le système d'IA est censé traiter.

Réception: les parties prenantes jugent que le système d'IA est fonctionnellement complet, qu'il offre un niveau de qualité acceptable et qu'il est prêt à être déployé.

Surveillance et examen du risque: il convient que les organismes passent en revue les résultats de la vérification, des essais et de la validation afin d'avoir connaissance des événements et des conditions entraînant des risques conformément au plan de traitement du risque (voir l'[ISO/IEC 23894](#)).

6.2.5 Déploiement

Le système d'IA est installé, déployé ou configuré pour fonctionner dans un environnement cible. Les processus de l'étape de déploiement peuvent inclure par exemple:

Cible: les systèmes d'IA peuvent être développés dans un environnement puis déployés dans un autre. Par exemple, un système de conduite autonome peut être développé dans un laboratoire, puis déployé dans des millions d'automobiles. D'autres types de systèmes d'IA peuvent être développés sur des dispositifs clients, puis déployés dans le nuage. Pour certains systèmes d'IA, il est important de faire la distinction entre les composants logiciels qui sont déployés et le modèle qui peut être déployé séparément et qui est utilisé par le logiciel à l'exécution. Le logiciel et le modèle peuvent être déployés indépendamment.

Traitement du risque: il convient que les organismes passent en revue et améliorent les processus et les contrôles de management du risque et éventuellement mettent à jour le plan de traitement du risque (voir l'ISO/IEC 23894).

6.2.6 Exploitation et suivi

Lors de l'étape d'exploitation et de suivi, le système d'IA fonctionne et est généralement disponible pour une utilisation.

Les processus qui peuvent se dérouler lors de l'étape d'exploitation et de suivi incluent par exemple:

Suivre: le système d'IA est suivi à la fois pour vérifier qu'il fonctionne normalement et pour détecter les incidents, y compris l'indisponibilité, les défaillances d'exécution ou les erreurs. Ces événements sont signalés aux fournisseurs d'IA pertinents afin qu'ils prennent des mesures.

Réparer: si le système d'IA connaît une défaillance ou subit des erreurs, il peut s'avérer nécessaire d'effectuer des réparations sur le système.

Mettre à jour: le logiciel, les modèles et le matériel du système d'IA peuvent être mis à jour afin de satisfaire à de nouvelles exigences et d'améliorer les performances et la fiabilité.

Fournir une assistance: les utilisateurs du système d'IA reçoivent toute l'assistance nécessaire pour utiliser le système avec succès.

Surveillance et examen du risque: il convient que les organismes surveillent les systèmes d'IA pendant l'exploitation afin de garantir et d'améliorer la qualité et l'efficacité du processus de management du risque (voir l'ISO/IEC 23894).

6.2.7 Validation continue

Si le système d'IA utilise l'apprentissage continu, l'étape d'exploitation et de suivi est prolongée par une étape supplémentaire de validation continue. À cette étape, un entraînement incrémentiel a lieu de façon continue tandis que le système fonctionne en production. Le bon fonctionnement du système d'IA est vérifié en permanence à l'aide de données d'essai. Il arrive également que les données d'essai elles-mêmes puissent nécessiter quelques mises à jour pour être plus représentatives des données de production actuelles et fournir ainsi une évaluation plus fidèle des capacités du système d'IA.

Amélioration continue du management du risque: il convient que la validation continue soit également utilisée pour permettre l'amélioration continue des processus de management du risque (voir l'ISO/IEC 23894).

6.2.8 Réévaluation

Après l'étape d'exploitation et de suivi, en fonction des résultats du travail du système d'IA, il peut apparaître nécessaire de procéder à une réévaluation. Les processus qui peuvent se dérouler durant l'étape de réévaluation incluent par exemple:

Évaluer les résultats d'exploitation: il convient que les résultats du système en exploitation soient évalués et appréciés par rapport aux objectifs et aux risques identifiés pour le système d'IA.

Affiner les objectifs: si les objectifs initiaux ne peuvent pas être atteints par le système d'IA, ou si les objectifs nécessitent d'être modifiés une fois acquise une expérience de l'exploitation du système. Cela amène à un affinement des objectifs.

Affiner les exigences: l'expérience d'exploitation peut fournir des preuves que certaines des exigences initiales ne sont d'une certaine manière pas valides, ce qui peut conduire à l'affinement des exigences, éventuellement par l'apparition de nouvelles exigences ou la suppression de certaines des exigences.

Surveillance et examen du risque: il convient que les organismes surveillent les événements et les conditions entraînant des risques tels que décrits dans le plan de traitement du risque (voir l'ISO/IEC 23894).

6.2.9 Mise hors service

À un moment donné, le système d'IA peut devenir obsolète dans la mesure où les réparations et les mises à jour ne suffisent pas à satisfaire aux nouvelles exigences. Les processus qui peuvent se dérouler durant l'étape de mise hors service incluent par exemple:

Déclassement et mise au rebut: si l'objectif du système d'IA n'existe plus, ou si une meilleure approche est apparue, le système d'IA peut être déclassé et intégralement mis au rebut. Cela peut inclure les données associées au système.

Remplacer: si l'objectif du système d'IA continue d'être pertinent, mais qu'une meilleure approche est apparue, le système d'IA (ou des composants du système d'IA) peut être remplacé.

7 Aperçu fonctionnel du système d'IA

7.1 Généralités

Le présent document définit le système d'IA comme un système technique qui génère des sorties telles que du contenu, des prévisions, des recommandations ou des décisions pour un ensemble donné d'objectifs définis par l'homme. Les systèmes d'IA ne comprennent pas; ils nécessitent des choix de conception, une ingénierie et une supervision de la part de l'homme. Le degré de supervision dépend du cas d'application. En règle générale, la supervision est au moins présente lors de l'entraînement et de la validation. Cette supervision est utile pour s'assurer que le système d'IA est développé et utilisé tel que prévu, et que les impacts sur les parties prenantes sont pris en compte de manière appropriée tout au long du cycle de vie du système.

La [Figure 5](#) représente une vue fonctionnelle d'un système d'IA, dans lequel les entrées sont traitées à l'aide d'un modèle pour produire des sorties, et ce modèle peut être construit directement ou à partir d'un apprentissage sur des données d'entraînement. Les parties représentées par des lignes en pointillé s'appliquent aux systèmes d'IA basés sur le ML.

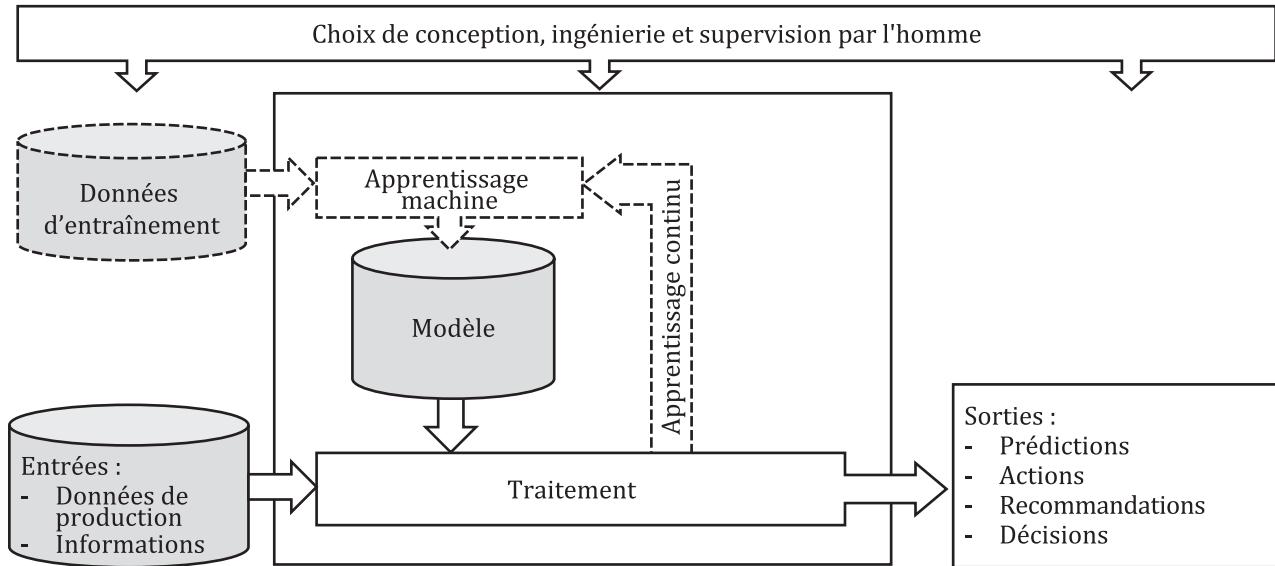


Figure 5 — Vue fonctionnelle d'un système d'IA

L'objectif de cette vue est de fournir une description non technique de ce que les systèmes d'IA font pour obtenir un résultat. En résumé, les systèmes d'IA contiennent un modèle qu'ils utilisent pour produire des prédictions et ces prédictions sont à leur tour utilisées pour formuler successivement des recommandations, des décisions et des actions, en tout ou en partie par le système lui-même ou par des êtres humains.

7.2 Données et informations

Les données peuvent être introduites dans le système d'IA en cours de production, auquel cas elles sont appelées données de production. Les données d'entrée peuvent nécessiter une préparation avant d'être présentées au système d'IA, comme l'extraction des caractéristiques pertinentes.

Les données d'entrée d'un système d'IA peuvent également être des informations plutôt que des données, généralement pour des tâches d'optimisation où la seule entrée nécessaire est l'information sur ce qui est à optimiser. Certains systèmes d'IA ne nécessitent aucune entrée, mais exécutent une tâche donnée sur demande (par exemple: la génération d'une image de synthèse).

Pour le ML, les données d'entraînement sont utilisées pour acquérir des informations au sujet du domaine d'intérêt et de la tâche à effectuer.

Les données ont d'autres utilisations pour le développement et l'évaluation des systèmes d'IA (voir [5.10](#)).

7.3 Connaissance et apprentissage

Le modèle utilisé par le système d'IA pour son traitement et pour la résolution de problèmes est une représentation des connaissances lisible par une machine.

Il existe deux types principaux de connaissances: les connaissances déclaratives et les connaissances procédurales:

- les connaissances déclaratives se rapportent à ce qui est. Elles sont faciles à verbaliser et se traduiront par des déclarations. Par exemple, «l'amanite phalloïde est un champignon vénéneux» est une connaissance déclarative;
- les connaissances procédurales concernent la manière de faire quelque chose. Elles sont souvent difficiles à verbaliser. Elles se traduiront par des procédures. Par exemple, pour savoir si un champignon est vénéneux, il est possible d'appliquer des connaissances procédurales: «Si vous

disposez d'un livre sur les champignons, consultez-le pour voir si vous pouvez identifier votre champignon. Si tel est le cas, le livre vous donnera la réponse. Si vous ne pouvez pas, allez voir un pharmacien».

Les connaissances ont différentes représentations possibles, implicites ou explicites.

Les connaissances peuvent également provenir de différentes sources, en fonction des algorithmes utilisés: elles peuvent préexister, elles peuvent être acquises par des processus de détection et d'apprentissage, ou elles peuvent être une combinaison des deux.

Systèmes heuristiques: les systèmes d'IA qui n'impliquent pas d'apprentissage sont appelés heuristiques. Les systèmes experts classiques ou les systèmes de raisonnement dotés d'une base de connaissances fixe en sont de bons exemples. Dans ces cas, les développeurs du système tirent parti des connaissances humaines pour fournir des règles raisonnables pour le comportement du système d'IA.

Systèmes d'IA basés sur le ML: il est dit des systèmes d'IA qui impliquent un apprentissage qu'ils sont basés sur l'apprentissage machine. L'apprentissage implique des analyses informatiques d'un ensemble de données d'entraînement dans le but de détecter des formes, de construire un modèle et de comparer la sortie du modèle obtenu aux comportements attendus. Il est également appelé entraînement. La base de connaissances obtenue est un modèle entraîné basé sur une fonction mathématique et un ensemble d'entraînement qui représente la meilleure approximation du comportement sur la base d'un environnement donné.

Apprentissage continu: les systèmes d'IA varient également en ce qui concerne le moment et la manière dont les données sont acquises. Dans certains cas, la base de connaissances est statique et fournie dès le départ, avec les composants préprogrammés du système. Dans d'autres cas, la base de connaissances change ou s'adapte au fil du temps, les informations étant mises à jour durant son exploitation. Les systèmes d'apprentissage machine peuvent être caractérisés par le moment, au cours de leur vie, où l'apprentissage se produit. Dans de nombreux cas, une phase d'entraînement initiale produit une certaine approximation de la fonction cible réelle, et le système continue tel quel sans mettre à jour cette représentation interne sur la base de nouveaux exemples. Une autre approche, appelée apprentissage continu ou apprentissage tout au long de la vie, étale l'apprentissage dans le temps; le modèle est mis à jour de façon itérative à mesure que de nouvelles données sont disponibles. Dans la pratique, les modèles qui utilisent l'apprentissage tout au long de la vie mettent généralement en œuvre une combinaison des deux approches; après une phase initiale d'entraînement au cours de laquelle l'essentiel de l'apprentissage se produit, le modèle est affiné au fil du temps avec de nouvelles données.

7.4 Des prédictions aux actions

7.4.1 Généralités

Le résultat du traitement des entrées par le système d'IA peut être de différentes natures, en fonction du niveau d'automatisation du système. Selon le cas d'application, le système d'IA peut produire uniquement un résultat technique brut (prédictions) ou il peut prendre des mesures plus efficaces et proposer ou réaliser lui-même des actions sur l'environnement (recommandations, décisions et enfin actions).

Dans le cas de la classification, les résultats erronés sont généralement classés en erreurs faussement positives ou faussement négatives. Un faux positif est décrit comme une prédition positive alors que le résultat réel est négatif. Un faux négatif est le résultat de la prédition incorrecte d'un résultat négatif par le modèle. Il est nécessaire que les utilisateurs de systèmes d'IA comprennent les effets d'un résultat erroné, y compris la possibilité d'une prédition biaisée. Ces problèmes peuvent refléter directement les caractéristiques des outils, des processus ou des données utilisés pour développer le système.

Un point essentiel est que les sorties de l'IA sont sujettes à erreur. Les sorties ont une probabilité d'être correctes, plutôt que d'être absolument vraies. Il est nécessaire que les concepteurs et les utilisateurs de systèmes d'IA comprennent que ces systèmes peuvent produire des sorties incorrectes et comprennent les implications de l'utilisation de ces sorties incorrectes en termes de redevabilité.

7.4.2 Prédiction

Le terme «prédiction» se rapporte au tout premier résultat d'un système d'IA.

Les systèmes d'IA font des prédictions en appliquant un modèle à de nouvelles données ou situations. Dans le scénario de crédit présenté en [7.4.3](#), un système d'IA a été développé à partir de dossiers de prêts antérieurs. Pour poursuivre l'exemple, lorsqu'une nouvelle personne demande un prêt, ses informations sont transmises au modèle, qui produit alors une estimation de la probabilité que cette personne rembourse le prêt.

NOTE Dans l'usage de l'intelligence artificielle, la prédiction n'implique pas nécessairement une déclaration au sujet de l'avenir - elle fait uniquement référence à la sortie d'un système d'IA, qui peut être un type de fleur dans une image, ou une traduction dans une autre langue.

7.4.3 Décision

Les décisions correspondent au choix d'une ligne de conduite spécifique, avec l'intention de l'appliquer.

Les décisions peuvent être prises soit par le système lui-même, soit par des êtres humains, sur la base des résultats du système. Elles peuvent être prises sur la base de recommandations ou se fonder directement sur des prédictions.

Par exemple, s'il est prévu qu'une personne présente un bon risque de crédit, un responsable de prêts humain peut analyser ce résultat avec les autres informations concernant cette personne et la situation du prêteur, puis décider d'approuver la demande de prêt de cette personne. Autrement, le système peut recommander lui-même d'accorder le prêt et estimer la probabilité qu'il s'agisse de la meilleure ligne de conduite en ce qui concerne les attentes du prêteur, de sorte qu'un responsable de prêts qui juge cette probabilité acceptable décide d'accorder le prêt. La demande de prêt peut également être approuvée automatiquement, en appliquant des seuils de décision du système à ces recommandations.

La supervision et le jugement humains sont impliqués de diverses manières dans ce processus de décision. Les seuils définis par l'homme sont généralement fixés en tenant compte des risques associés à l'automatisation des décisions. Même lorsque les décisions sont entièrement automatisées, les humains peuvent utiliser des prédictions pour surveiller les décisions obtenues.

7.4.4 Action

Les actions suivent les décisions, c'est le moment où les résultats du système d'IA commencent à avoir une incidence sur le monde réel (physique ou virtuel).

Réaliser une action est l'étape finale de l'application des informations dans un système d'IA. Par exemple, dans l'exemple de la demande de crédit en [7.4.2](#), une fois que le prêt de la personne est approuvé, les actions peuvent inclure la préparation des documents de prêt, l'obtention des signatures et l'émission des paiements. Prendre le cas d'un robot. Une action peut être une série d'instructions données aux actionneurs du robot pour qu'ils positionnent ses bras et ses mains. En fonction du système d'IA, l'action peut se dérouler à l'intérieur ou à l'extérieur des frontières du système d'IA.

8 Écosystème de l'IA

8.1 Généralités

La [Figure 6](#) représente un écosystème d'IA en termes de couches fonctionnelles. Les grands systèmes d'IA ne reposent pas sur une seule technologie, mais plutôt sur un mélange de technologies développées au fil du temps. Ces systèmes peuvent utiliser différentes technologies simultanément, par exemple les réseaux neuronaux, les modèles symboliques et le raisonnement probabiliste.

Chaque couche de la [Figure 6](#) utilise les ressources des couches inférieures pour la mise en œuvre de ses fonctions. Les cases plus claires sont des sous-composants d'une couche ou d'une fonction. La taille des couches et des sous-composants n'est pas indicative de leur importance.

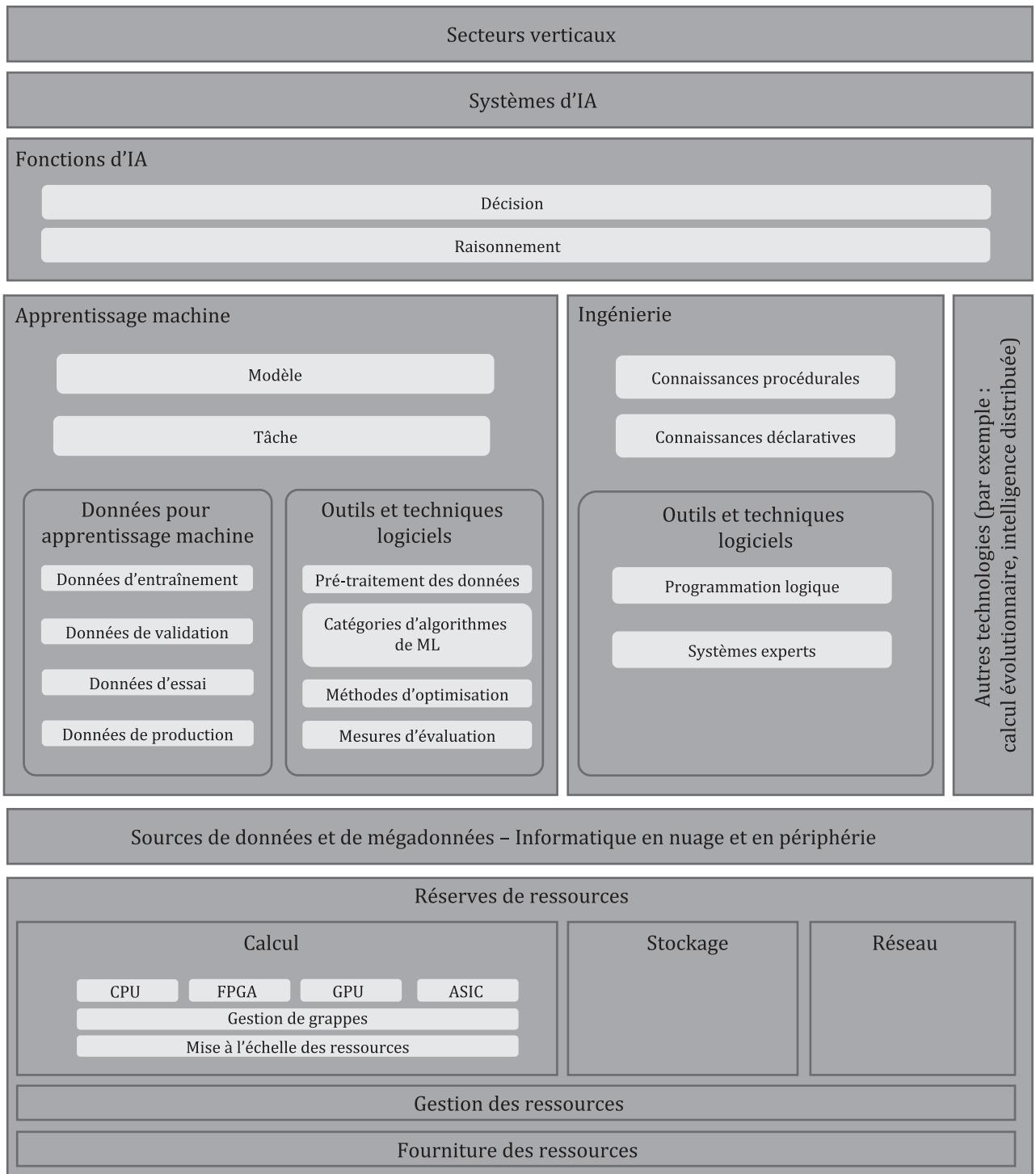


Figure 6 — Écosystème de l'IA

La création de systèmes d'IA reste un sujet de recherche permanent. Pendant ce temps-là, l'utilisation de la technologie de l'IA devient une partie inhérente de nombreux secteurs, chacun ayant ses besoins, ses valeurs et ses contraintes légales propres.

Les applications d'IA spécialisées, telles que celles utilisées pour la vision artificielle ou le traitement du langage naturel, deviennent elles-mêmes les éléments constitutifs de la mise en œuvre de différents produits et services. Ces applications permettent la conception de systèmes d'IA spécialisés et, par conséquent, définissent les priorités en matière de recherche et de développement.

La technologie de l'IA nécessite souvent l'utilisation d'importantes ressources de calcul, de stockage et de réseau, par exemple pendant la phase d'entraînement d'un système d'apprentissage machine. Ces ressources, telles que représentées à la [Figure 6](#), peuvent être fournies efficacement grâce à l'informatique en nuage.

Les paragraphes suivants décrivent les principaux composants de l'écosystème de l'IA tel qu'illustré à la [Figure 6](#).

8.2 Systèmes d'IA

Les systèmes d'IA peuvent être utilisés dans de nombreuses applications et pour résoudre une multitude de tâches. [L'Article 9](#) décrit des exemples d'applications qui utilisent l'IA, telles que la reconnaissance d'image, le traitement du langage naturel et la maintenance prédictive. [L'Article 5](#) énumère de nombreuses catégories de tâches que les systèmes d'IA peuvent couvrir.

Les systèmes d'IA suivent un chemin fonctionnel global où les informations sont acquises, soit par codage en dur (par l'ingénierie), soit par apprentissage machine afin de construire un modèle du domaine. Les informations, codées sous la forme d'un modèle, sont ensuite appliquées à un niveau de raisonnement, où des solutions potentielles sont calculées, puis à un niveau de décision, où un choix est fait parmi les actions potentielles pouvant permettre d'atteindre l'objectif. Le niveau de raisonnement comprend le raisonnement spatial, le raisonnement temporel, le raisonnement de bon sens, l'application d'une politique calculée ou toute forme de raisonnement pouvant être codée. Le niveau de décision inclut le choix basé sur les préférences ou l'utilité des actions.

8.3 Fonction d'IA

Une fois le modèle construit, une fonction d'IA a pour rôle de calculer une prédiction, une recommandation ou plus généralement une décision qui aiderait à atteindre l'objectif actuel du système d'IA.

Le raisonnement consiste uniquement à appliquer les données disponibles dans la situation actuelle au modèle et à lui demander quelles sont les options possibles.

Les technologies qui mettent en œuvre des formes de raisonnement incluent par exemple la planification, le raisonnement bayésien, les démonstrateurs automatiques de théorèmes, le raisonnement temporel et spatial et les systèmes de raisonnement ontologiques.

Parmi ces options possibles qui permettraient probablement d'atteindre l'objectif, le système a encore besoin de décider laquelle est la meilleure.

Les préférences et les utilités entrent en jeu: un taxi automatisé maximisera le bien-être du client, un programme de jeu de poker maximisera son profit.

8.4 Apprentissage machine

8.4.1 Généralités

L'apprentissage machine est un processus qui utilise des techniques de calcul pour permettre aux systèmes d'apprendre à partir de données ou de l'expérience. Il utilise un ensemble de méthodes statistiques pour trouver des formes dans les données existantes et les utiliser ensuite pour faire des prédictions sur les données de production.

Dans la programmation informatique traditionnelle, un programmeur spécifie la logique permettant de résoudre un problème donné en spécifiant les étapes exactes de calcul à l'aide d'un langage de programmation. À l'inverse, la logique d'un modèle d'apprentissage machine dépend en partie des données utilisées pour entraîner le modèle. Ainsi, les calculs, ou étapes, nécessaires pour résoudre le problème ne sont pas déterminés au préalable.

En outre, contrairement à la programmation informatique traditionnelle, les modèles d'apprentissage machine peuvent s'améliorer au fil du temps sans être réécrits, en étant ré-entraînés sur de nouvelles données supplémentaires et en utilisant des techniques pour optimiser les paramètres du modèle et les caractéristiques des données.

8.5 Ingénierie

8.5.1 Généralités

Dans les approches d'ingénierie par des experts humains, le traitement repose exclusivement sur l'expertise du développeur et sa compréhension de la tâche. Les connaissances ne sont pas acquises à partir de données, mais par codage en dur effectué par le développeur sur la base de son expérience dans un domaine spécifique.

Il existe deux types principaux de connaissances: les connaissances déclaratives et les connaissances procédurales. Voir [7.3](#) pour plus de détails sur les deux types de connaissances.

8.5.2 Systèmes experts

Comme le nom l'indique, un système expert est un système d'IA qui encapsule les connaissances fournies par un expert humain dans un domaine spécifique afin de déduire des solutions à des problèmes.

Un système expert se compose d'une base de connaissances, d'un moteur inférence et d'une interface utilisateur. La base de connaissances stocke les connaissances déclaratives d'un domaine spécifique, qui englobent à la fois des informations factuelles et heuristiques. Le moteur inférence contient des connaissances procédurales: l'ensemble des règles et la méthodologie du raisonnement. Il associe les faits fournis par l'utilisateur et les informations de la base de connaissances.

L'inférence s'effectue à l'aide de règles prédéfinies en fonction de l'expert et avec des évaluations d'énoncés logiques. Les classes de problèmes qui peuvent être résolus à l'aide de systèmes experts incluent la classification, le diagnostic, la surveillance et la prédiction.

8.5.3 Programmation logique

La programmation logique est une forme de programmation basée sur des langages de programmation qui expriment une logique formelle. Prolog est un exemple de langage de programmation logique.

Les logiques formelles pour l'IA ont fait l'objet de nombreuses recherches. De nombreux types de logiques formelles visent à modéliser le raisonnement humain dans diverses situations. La programmation logique fournit un cadre pour mettre en œuvre ces modèles de raisonnement humain. Il est nécessaire que les agents d'IA soient capables de reproduire différents types de raisonnement d'une manière clairement spécifiée, transparente et explicable.

La programmation logique avec des énoncés déclaratifs, associée à un traitement du langage naturel puissant, peut amener un agent à raisonner par analogie, à tirer des conclusions et à généraliser à propos des objets et de l'environnement.

EXEMPLE Apache Jena^[19] est un cadre Web sémantique qui fournit un moteur inférence.

8.6 Sources de données et de mégadonnées — informatique en nuage et en périphérie

8.6.1 Sources de données et de mégadonnées

Tous les systèmes de ML utilisent des données. Ces données peuvent prendre des formes diverses. Dans certains cas, les données utilisées par les systèmes de ML sont des mégadonnées. Ce bloc à la [Figure 6](#) représente les sources, les formats et le traitement type des mégadonnées indépendamment de leurs usages. Le présent paragraphe décrit plus en détail les composants principaux représentés à la [Figure 7](#).

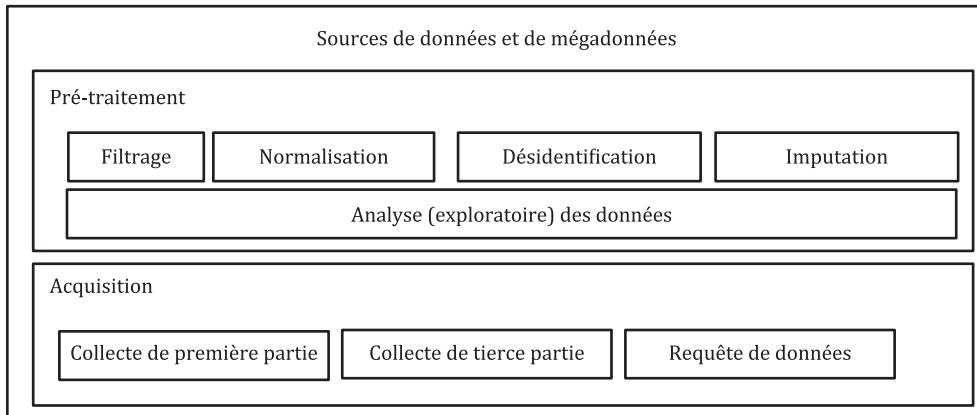


Figure 7 — Sources de données et de mégadonnées

Les mégadonnées sont des ensembles de données étendus dont les caractéristiques en termes de volume, de variété, de vitesse et de variabilité nécessitent des technologies et des techniques spécialisées pour les traiter et les valoriser. Par exemple, des technologies ont été développées spécifiquement pour permettre le traitement distribué de grands ensembles de données à l'aide de grappes d'ordinateurs tout en utilisant des modèles de programmation simples. De plus, des technologies de stockage et de base de données ont été développées spécifiquement pour gérer de grands volumes de données qui peuvent être composés d'autres grands volumes.

Les mégadonnées sont devenues importantes car les organismes ont augmenté l'ampleur et la profondeur de la collecte de données et nécessitent donc des technologies et des techniques spécialisées pour en extraire des informations.

Pour plus d'informations sur les mégadonnées, voir l'ISO/IEC 20546 et l'ISO/IEC 20547-3.

Les mégadonnées ont de nombreuses utilisations pour les systèmes d'IA, et elles sont un facilitateur de beaucoup de ces systèmes. La disponibilité de grandes collections de données non structurées dans différents domaines d'application fournit de nouvelles informations grâce à l'utilisation de techniques d'IA telles que la découverte de connaissances et la reconnaissance des formes. La disponibilité de grandes quantités de données pour l'entraînement permet une amélioration des modèles d'apprentissage machine capables d'être utilisés dans un large éventail d'applications.

Les données peuvent être acquises par le même organisme que celui qui les utilise (collecte par première partie). Par exemple, les détaillants utilisent les données de transaction qu'ils collectent à partir des systèmes de point de vente qu'ils possèdent. Les données peuvent également être acquises par des tiers, tels que des organismes de recherche et d'autres fournisseurs de données qui collectent des données puis les vendent ou les partagent avec d'autres organismes qui les utilisent. De plus, les données peuvent être acquises en lançant des requêtes et en combinant des données provenant de différents ensembles de données, d'une première comme d'une tierce partie.

Les données peuvent provenir de nombreuses sources telles que:

- point de vente et autres transactions;
- sondages ou enquêtes;
- recherches statistiques;
- observations enregistrées;
- capteurs;
- images;
- enregistrements audio;

- documents;
- interactions avec des systèmes.

8.6.2 Informatique en nuage et en périphérie

L'informatique en nuage est un paradigme destiné à permettre l'accès en réseau à un regroupement évolutif et élastique de ressources physiques ou virtuelles partageables, avec un provisionnement et une administration libres à la demande, voir l'ISO/IEC 17788 et l'ISO/IEC 17789.

L'informatique en nuage est généralement associé à de grands centres de données centralisées, capables de fournir de très grandes capacités de traitement et de stockage de données. Ces grandes capacités peuvent être essentielles pour certaines parties du cycle de vie de l'IA, en particulier lors du traitement de grands ensembles de données destinés à entraîner les systèmes d'IA et à construire les modèles utilisés en leur sein.

L'informatique en périphérie est de l'informatique distribuée dans laquelle le traitement et le stockage des données s'effectuent à la périphérie du réseau ou à proximité, la proximité étant définie par les exigences du système. La périphérie est la frontière entre les entités numériques et physiques pertinentes, délimitée par des capteurs et des actionneurs en réseau (voir l'ISO/IEC 23188).

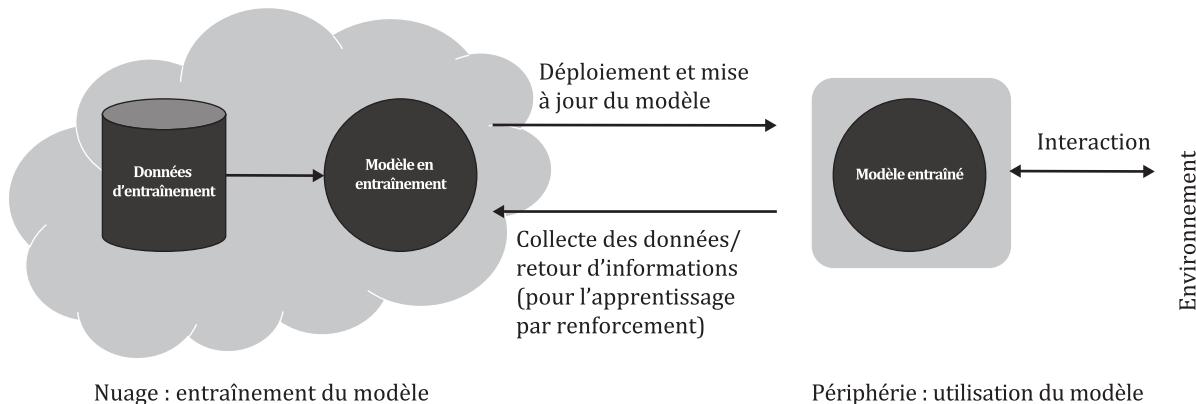
L'informatique en périphérie concerne en grande partie le placement et le fonctionnement des composants logiciels et le stockage des données. Lorsque des composants logiciels, tels que ceux associés aux systèmes d'IA, traitent avec des dispositifs de l'IdO (capteurs et actionneurs), il est souvent nécessaire de réduire au minimum les latences et de produire des résultats avec des contraintes de temps importantes (souvent appelées temps réel), ou il existe souvent un besoin de résilience afin qu'un système puisse continuer à fonctionner si les communications sont interrompues, ou il existe un besoin de protéger la confidentialité des données personnelles capturées à partir de dispositifs en périphérie. Pour y parvenir, il peut être nécessaire que le traitement et le stockage des données soient effectués en périphérie ou à proximité. Pour plus de détails, voir l'ISO/IEC 23188.

Toutefois, il est important de comprendre que l'informatique en nuage peut être déployée dans de nombreux endroits dans un environnement informatique distribué, y compris dans des endroits qui ne sont pas centralisés et qui sont proches de la périphérie. Sous cette forme, l'informatique en nuage peut offrir un déploiement souple et dynamique des logiciels comme des données, en utilisant un traitement et un stockage de données virtualisés, associés à une mise en commun des ressources et à une élasticité et une évolutivité rapides, afin de permettre un placement et un fonctionnement appropriés des composants des systèmes d'IA.

Il arrive souvent que les systèmes d'informatique en périphérie soient associés à des systèmes centralisés afin de créer des solutions complètes, qui tirent parti des capacités de chaque type de système.

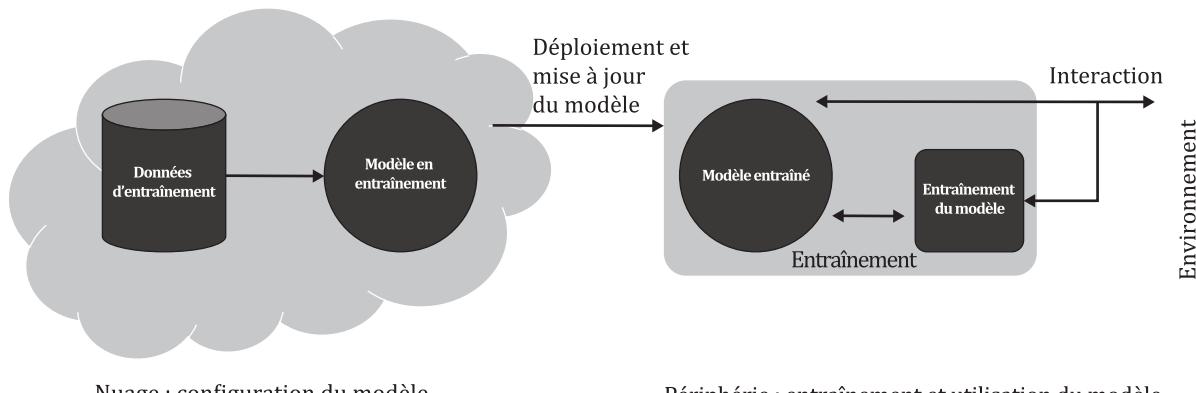
Trois conceptions principales de systèmes de ML associent le nuage et la périphérie: l'entraînement de modèle dans le nuage, l'entraînement de modèle en périphérie et l'entraînement de modèle dans le nuage et en périphérie.

- a) Les services en nuage peuvent être utilisés comme plateforme centralisée pour l'entraînement des modèles de ML ([Figure 8](#)). En raison des restrictions en matière de ressources des dispositifs en périphérie, les tâches à forte intensité de calcul et de stockage liées à l'entraînement, à la validation et à la maintenance des modèles sont effectuées à l'aide d'une infrastructure en nuage. Une fois entraîné, le modèle est déployé, appliqué et, si nécessaire, mis à jour sur des dispositifs en périphérie. Les données provenant des dispositifs en périphérie peuvent être utilisées pour effectuer des activités d'entraînement ou, comme dans le cas de l'apprentissage par renforcement, pour fournir des retours d'information sur la qualité du modèle.

**Figure 8 — Exemple d'entraînement d'un modèle dans le nuage**

Des exemples d'applications utilisant cette conception sont la détection d'attaque sur les routeurs de périphérie (pare-feux intelligents), la détection et la prévention des pannes dans les applications de contrôle industriel (maintenance proactive) et la reconnaissance des panneaux de signalisation pour les voitures autonomes.

- b) Si une approche centralisée n'est pas optimale pour l'entraînement de modèles personnalisés ou de modèles utilisés dans des contextes d'application spécifiques, un schéma différent ([Figure 9](#)) peut être appliqué. Il est basé sur l'entraînement de modèle effectué directement sur les dispositifs en périphérie (à condition que des ressources suffisantes soient disponibles sur ces dispositifs).

**Figure 9 — Exemple d'entraînement de modèle en périphérie**

Dans cette conception, seul un modèle initial (générique) est configuré et entraîné dans l'environnement en nuage. L'entraînement contextualisé ou personnalisé est effectué en périphérie, à l'aide de données réelles. Ce type d'entraînement de modèle convient le mieux aux systèmes entièrement automatiques qui utilisent des approches d'apprentissage non supervisé ou d'apprentissage par renforcement.

- c) Une approche hybride procède à l'entraînement de modèle à la fois dans le nuage et en périphérie ([Figure 10](#)). Cela peut s'avérer nécessaire lorsque la conception du système contient des dispositifs en périphérie. Dans certains cas, les services en nuage préparent le modèle entraîné initial qui est ensuite déployé en périphérie. Dans d'autres cas, les systèmes en périphérie entraînent leurs modèles locaux sur la base de leurs données locales et ne se transmettent pas les données entre eux ni ne les transmettent aux services en nuage. Les services en nuage peuvent également agir comme serveur de paramètres pour synchroniser les mises à jour des modèles des différents systèmes en périphérie, puis renvoyer les mises à jour synchronisées des modèles aux systèmes en périphérie pour qu'ils mettent à jour leurs propres modèles.

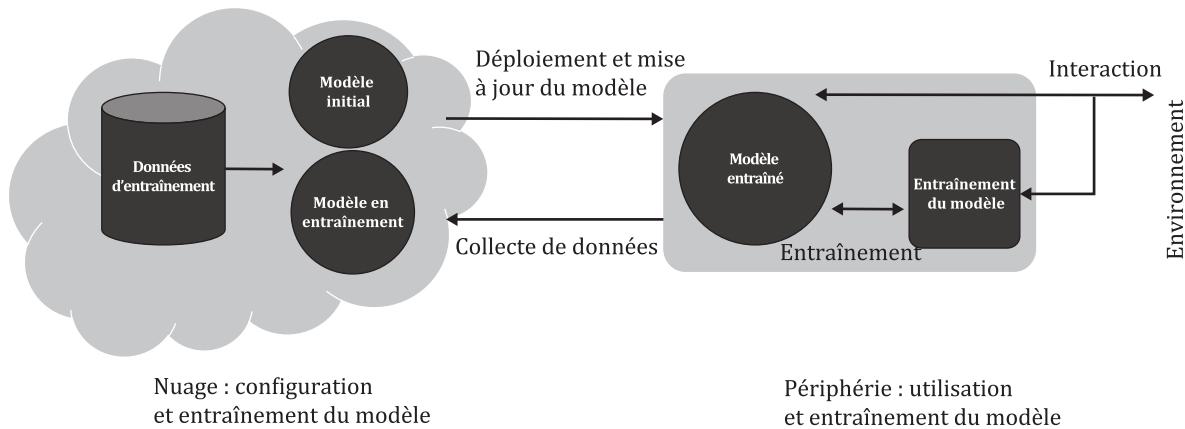


Figure 10 — Exemple d'entraînement de modèle dans le nuage et en périphérie

Cet exemple est appliquée pour un service de données basé sur une zone locale (par exemple: un service de collecte d'images par drone dans différents secteurs et un service d'appareils ménagers) qui permet d'améliorer la qualité du service grâce à des modèles d'entraînement mis à jour plutôt qu'aux modèles entraînés initiaux.

Une autre approche hybride comprenant le téléchargement de modèles peut être envisagée. Pour cette méthode, un modèle entraîné par un système en périphérie est envoyé à un référentiel de modèles dans l'environnement en nuage et est distribué à d'autres systèmes en périphérie fonctionnant dans le même environnement ou dans un environnement similaire si le modèle entraîné est plus performant que le précédent. Cette approche peut être appliquée aux techniques d'apprentissage par transfert et de compression de modèle. Pour un exemple d'apprentissage par transfert, un modèle entraîné pour l'identification des numéros de maison dans la vue d'une rue peut être utilisé pour reconnaître des numéros manuscrits. Un modèle initial ou un modèle déjà entraîné à la résolution d'un problème spécifique peut être appliqué à des problèmes similaires. Les techniques de compression de modèle peuvent également être utilisées pour un dispositif en périphérie disposant d'une puissance de calcul inférieure. Un modèle entièrement entraîné dans le service en nuage disposant de ressources de calcul abondantes peut être compressé avant d'être utilisé dans un système en périphérie disposant de moins de ressources.

8.7 Réserves de ressources

8.7.1 Généralités

La [Figure 6](#) montre les ressources requises pour soutenir l'écosystème de l'IA. Les ressources telles que le calcul, la mise en réseau et le stockage sont essentielles pour soutenir les systèmes d'IA.

Le développement et le déploiement des systèmes d'IA peuvent se produire sur des ressources de différentes tailles, allant des services en nuage centralisés et des centres de données hors nuage aux serveurs (ou grappes de serveurs), systèmes d'informatique en périphérie, dispositifs mobiles et dispositifs de l'IdO. Certains de ces systèmes peuvent être limités au niveau des ressources, en termes de capacité de traitement, de capacité de stockage des données, de bande passante et de latence réseau. Cela est particulièrement le cas pour les systèmes et les dispositifs en périphérie. Les ressources de traitement pour les systèmes d'IA peuvent être tout arrangement et toute combinaison de GPU, NPU, CPU et autres types de processeurs uniques ou multiples dans des systèmes uniques ou multiples qui peuvent être organisés en grappes de calcul.

Les exigences d'un système d'IA en matière de ressources de calcul peuvent varier en fonction de l'utilisation de l'apprentissage machine ou de l'apprentissage profond, ainsi que des types de charge de travail (par exemple: entraînement et inférence avec diverses topologies). Par conséquent, des solutions de calcul hétérogènes peuvent être nécessaires selon les cas en fonction de la charge de travail et du système d'IA considérés. Par exemple, des accélérateurs matériels (GPU, NPU, FPGA, DSP, ASIC, etc.)

peuvent être utilisés pour les charges de travail d'IA à forte intensité de calcul en calcul, telles que l'entraînement de certaines topologies de réseaux neuronaux.

Pour satisfaire aux besoins des différents systèmes d'IA, il est nécessaire que le provisionnement des ressources permette une gestion automatisée, y compris le provisionnement à la demande et la coordination des ressources hétérogènes (par exemple: provisionnement de ressources sur site, dans le nuage et en périphérie).

8.7.2 Circuit intégré spécifique

Un ASIC est un type de circuit intégré personnalisé pour une utilisation particulière. Les ASIC sont une option pour la fourniture de fonctionnalités spécifiques à l'IA.

Un ASIC peut être personnalisé en tant qu'accélérateur afin d'accélérer le processus d'IA en fournissant des fonctions telles que des blocs multiplieurs-accumulateurs parallèles dédiés, une allocation de mémoire optimisée et une arithmétique de faible précision. Un ASIC peut également être configuré comme coprocesseur pour fournir des fonctions de pré- et post-traitement des données pour les tâches d'IA telles que le recadrage et le redimensionnement d'images, la transformation, la réduction du bruit et la fusion de données de détection.

Par rapport aux processeurs génériques (par exemple: les CPU et les GPU), les ASIC sont généralement conçus, produits et utilisables uniquement pour des scénarios spécifiques tels que la mise en œuvre de structures de réseaux neuronaux spécifiques. Les ASIC offrent une plus grande capacité de calcul pour l'IA avec des volumes spatiaux, des coûts et une consommation d'énergie moindres.

Les ASIC permettent de mettre en œuvre l'IA dans des dispositifs limités en termes d'espace et de puissance, tels que les téléphones portables. Les ASIC permettent également d'utiliser l'IA dans les dispositifs de l'IdO utilisés dans divers domaines tels que la production industrielle, la médecine, l'assainissement, la sécurité et les scénarios de maison intelligente.

9 Domaines de l'IA

9.1 Vision artificielle et reconnaissance d'image

Le présent document définit la vision artificielle comme la «capacité d'une unité fonctionnelle à acquérir, traiter et interpréter des données représentant des images ou une vidéo» ([3.7.1](#)). La vision artificielle est étroitement liée à la reconnaissance d'image, par exemple au traitement des images numériques. Les données visuelles proviennent généralement d'un capteur d'image numérique, d'une image analogique numérisée ou d'un autre dispositif d'entrée graphique. Aux fins du présent document, les images numériques incluent les variantes tant fixes que mobiles.

Les images numériques existent sous la forme d'une matrice de nombres qui représentent les échelles de gris ou les couleurs de l'image capturée ou, dans d'autres cas, une collection de vecteurs. Les images numériques peuvent inclure des métadonnées qui décrivent les caractéristiques et les attributs associés à l'image. Les images numériques peuvent être compressées pour économiser de l'espace de stockage et améliorer leurs performances de transmission sur les réseaux numériques.

Ci-après des exemples d'applications de l'IA basées sur la vision artificielle et la reconnaissance d'image:

- identification d'images spécifiques dans un ensemble d'images (par exemple: des images de chiens dans un ensemble d'images d'animaux);
- véhicules autonomes: détection et identification des feux de signalisation et des objets dans les véhicules automatisés;
- diagnostics médicaux: détection de maladies et d'anomalies dans l'imagerie médicale;
- contrôle de la qualité (par exemple: repérage des pièces défectueuses sur une chaîne d'assemblage);
- reconnaissance faciale.

Les tâches fondamentales de la vision artificielle incluent l'acquisition d'images, le rééchantillonnage, la mise à l'échelle, la réduction du bruit, l'amélioration du contraste, l'extraction de caractéristiques, la segmentation, la détection d'objets et la classification.

Plusieurs méthodes existent pour accomplir les tâches de vision artificielle dans les systèmes d'IA. L'utilisation de réseaux neuronaux convolutionnels profonds ([5.12.1.4](#)) a gagné en popularité ces dernières années en raison de leur grande exactitude dans les tâches de classification des images et de leurs performances en matière d'entraînement et de prédiction.

9.2 Traitement du langage naturel

9.2.1 Généralités

Le traitement du langage naturel est un traitement de l'information basé sur la compréhension du langage naturel et la génération automatique de texte. Il englobe l'analyse et la génération de langage naturel, avec du texte ou de la parole. En utilisant les capacités du NLP, les ordinateurs peuvent analyser des textes écrits en langage humain et identifier des concepts, des entités, des mots-clés, des relations, des émotions, des sentiments et d'autres caractéristiques, ce qui permet aux utilisateurs d'extraire des informations du contenu. Grâce à ces capacités, les ordinateurs peuvent également générer du texte ou de la parole pour communiquer avec les utilisateurs. Tout système qui prend du langage naturel comme entrée ou comme sortie, sous forme de texte ou de parole, et qui est capable de le traiter, utilise des composants de traitement du langage naturel. Un exemple d'un tel système est un système de réservation automatisé pour une compagnie aérienne, capable de prendre les appels des clients et de réserver des vols pour eux. Un tel système a besoin d'un composant de compréhension du langage naturel et d'un composant de génération automatique de texte.

Ci-après des exemples d'applications de l'IA basées sur le traitement du langage naturel:

- la reconnaissance de l'écriture manuscrite (par exemple: conversion de notes manuscrites sous forme numérique);
- la reconnaissance de la parole (par exemple: compréhension du sens de paroles humaines);
- la détection des courriers indésirables (par exemple: utilisation du sens des mots dans un message électronique pour déterminer si ce message est classé comme indésirable);
- les assistants numériques personnels et les agents conversationnels en ligne qui peuvent utiliser la compréhension du langage naturel et la génération automatique de texte (y compris la reconnaissance et la génération de la parole) pour offrir des interfaces utilisateur conversationnelles;
- le résumé;
- la génération de texte;
- la recherche de contenu.

Le NLP est également utilisé dans de nombreux systèmes applicatifs tels que les agents conversationnels, les systèmes de publicité basés sur le contenu, les systèmes de traduction vocale et les systèmes d'apprentissage en ligne.

9.2.2 Composants de traitement du langage naturel

9.2.2.1 Généralités

Les composants de NLP accomplissent différentes tâches. Les tâches les plus courantes sont:

NLU: un composant de NLU convertira du texte ou un discours en une description interne qui est censée contenir la sémantique de l'entrée. La difficulté vient de l'ambiguïté inhérente aux langages naturels: les mots et les phrases étant par nature polysémiques, le résultat d'un NLU est donc sujet à l'erreur.

NLG: un composant de NLG convertira une description interne en texte ou en discours compréhensible par un être humain. Cette tâche peut impliquer d'ajuster la formulation afin qu'elle semble plus naturelle à l'utilisateur.

POS: un composant d'étiquetage morpho-syntaxique (POS) est utilisé pour catégoriser chaque mot de l'entrée en tant qu'objet grammatical: s'il s'agit d'un nom, d'un adjectif, d'un verbe, etc. L'étiquetage morpho-syntaxique (POS) est également affecté par la polysémie.

NER: un composant de NER cherchera à reconnaître et à étiqueter les noms dénotatifs d'une personne, d'un lieu, d'un organisme ou d'une autre entité pour des séquences de mots dans un flux de texte ou de paroles. Selon l'entité, des informations supplémentaires peuvent être extraites. Par exemple, pour les personnes, leur titre ou leur fonction est utile.

Réponse aux questions: un composant de réponse aux questions essaiera de donner la réponse la plus appropriée à une question humaine. L'utilisateur pose une question en langage naturel et le système fournit une réponse en langage naturel.

TA: un composant de TA traduit automatiquement un contenu en langage naturel d'une langue à une autre. Cela peut s'effectuer de texte à texte, de paroles à texte, de paroles à paroles ou de texte à paroles. La difficulté vient de la polysémie, dans le cadre de laquelle un mot a plusieurs significations, ainsi que d'autres sources, telles que les références entre ou à l'intérieur des phrases ou les intentions non dites. Dans de nombreux cas, plusieurs traductions sont possibles.

OCR: un composant d'OCR cherchera à convertir des documents écrits sous forme d'images (éventuellement numérisées) en une description codée numérique de leur contenu: texte, tableaux, figures, titres et leurs relations.

Extraction de relations: un composant d'extraction de relations gère la tâche d'extraction de relations entre des entités nommées ou même entre n'importe quelles entités de l'entrée. Par exemple, le composant peut identifier dans un texte d'entrée au sujet de films qu'«Al Pacino» a «joué» dans le film «Serpico».

IR: un composant d'IR ou de recherche cherchera à répondre aux besoins d'information de l'utilisateur en effectuant des recherches dans une collection de contenu non structuré. La requête de l'utilisateur exprimant son besoin en informations est comparée algorithmiquement à chaque élément de la collection, afin de prédire sa pertinence par rapport au besoin en informations de l'utilisateur. La sortie de ce composant est généralement présentée à l'utilisateur sous la forme d'une liste d'éléments sélectionnés, classés par ordre décroissant de pertinence. Les composants de recherche documentaire peuvent être développés pour un large éventail de types d'éléments d'information différents, y compris le texte libre, les documents semi-structurés, les documents structurés, l'audio, l'image et la vidéo, et ce dans différents langages naturels.

Analyse de sentiments: un composant d'analyse de sentiments cherchera à identifier et à catégoriser informatiquement les opinions exprimées dans un texte, un discours ou une image. Cette analyse est également appelée «fouille d'opinions». Les aspects subjectifs peuvent par exemple inclure des sentiments positifs ou négatifs.

Résumé automatique: un composant de résumé automatique exprimera de manière plus concise les informations importantes d'un contenu en recourant à l'une des deux approches qui existent (ou une combinaison de celles-ci). Le résumé extractif qui sélectionne le contenu clé pertinent du contenu source pour produire une version résumée réduite. Le résumé abstractif qui cherche à synthétiser un nouveau texte plus court qui exprime les informations pertinentes. Le résumé abstractif est lié à la génération automatique de texte.

Gestion du dialogue: un composant de gestion du dialogue aidera à gérer une série d'interactions entre un utilisateur et un système dans le but d'améliorer l'expérience de l'utilisateur en imitant une conversation en langage naturel. La gestion du dialogue utilise une palette d'approches, y compris des règles déclaratives spécifiant des réponses pour des déclencheurs d'entrée spécifiques et des approches basées sur l'apprentissage machine. La gestion du dialogue peut alimenter des interactions textuelles, par exemple pour fournir une expérience plus conversationnelle avec des composants de

réponse aux questions ou être intégrée à des composants de reconnaissance et de synthèse de la parole pour soutenir des applications dans les assistants personnels, les agents de service client en ligne ou la robotique de soins personnels.

9.2.2.2 Traduction automatique

La traduction automatique est une tâche de NLP dans laquelle un système informatique est utilisé pour traduire automatiquement un texte ou un discours d'un langage naturel à un autre.

En général, le processus de traduction par un humain se déroule en deux étapes. La première étape consiste à décoder le sens de la langue source. La seconde étape consiste à re-coder le sens dans la langue cible. Ce processus nécessite une connaissance approfondie de la grammaire, de la sémantique, de la syntaxe, des expressions idiomatiques, du contexte culturel et d'autres domaines.

Les défis techniques de la traduction automatique comprennent les sens multiples des mots, les considérations contextuelles, les différences grammaticales et les langues utilisant des systèmes d'écriture basés sur des logogrammes. Il existe de nombreuses approches en matière de traduction automatique, telles que les approches basées sur des règles, basées sur des exemples, statistiques, neuronales ou une combinaison de ces approches.

Ces dernières années, les réseaux neuronaux ont été utilisés pour effectuer des traductions automatiques, ce qui a permis des progrès considérables dans la fluidité et l'exactitude des traductions. Grâce à l'apprentissage profond, le modèle peut être entraîné et personnalisé pour des expressions spécifiques à un domaine afin d'atteindre des niveaux d'exactitude élevés.

9.2.2.3 Synthèse de la parole

Un système qui convertit un texte en langage naturel en parole est appelé système de conversion texte-parole.

En général, le processus de TTS comporte trois étapes: 1) analyse, 2) modélisation et 3) synthèse. Le naturel et l'intelligibilité sont des caractéristiques importantes d'un système de TTS. Le naturel décrit dans quelle mesure la sortie ressemble à la parole humaine, tandis que l'intelligibilité se rapporte à la facilité avec laquelle la sortie est comprise par les humains. Les systèmes de synthèse de la parole s'efforcent généralement de maximiser ces deux caractéristiques.

Diverses approches sont utilisées en synthèse de la parole, y compris la synthèse segmentale, la synthèse par formants, la synthèse articulatoire, la synthèse basée sur le HMM, la synthèse sinusoïdale et les DNN. Chaque approche a ses forces et faiblesses propres. Certains synthétiseurs de parole basés sur les DNN se rapprochent de la qualité de la voix humaine.

9.2.2.4 Reconnaissance de la parole

Le présent document définit la reconnaissance de la parole comme étant la conversion, par une unité fonctionnelle, d'un signal de parole en une représentation du contenu de la parole. La parole numérisée est une forme de données séquentielles, de sorte que les techniques qui peuvent traiter des données associées à un intervalle de temps peuvent être utilisées pour traiter les phonèmes de la parole.

Plusieurs approches qui utilisent des réseaux neuronaux ont été utilisées pour la reconnaissance de la parole. L'une d'entre elles implique l'utilisation de LSTM^[18]. Cette méthode permet d'entraîner et de déployer un réseau neuronal comme solution de reconnaissance de la parole sans qu'il soit associé à d'autres processus tels que les HMM, ce qui permet d'obtenir des performances de reconnaissance raisonnables.

Ci-après des exemples d'applications de l'IA basées sur la reconnaissance de la parole:

- systèmes de commande vocale;
- dictée numérique;

- assistants personnels.

9.2.2.5 Réponse aux questions

Les systèmes de réponse aux questions peuvent ingérer un grand nombre de pages de texte et appliquer la technologie de réponse aux questions pour répondre aux questions posées par les humains en langage naturel. Cette approche permet aux utilisateurs de «demander» et d'obtenir des réponses quasi instantanées à des questions complexes. Associée à d'autres API et à l'analyse avancée, la technologie de réponse aux questions se distingue de la recherche conventionnelle (qui est déclenchée par des mots-clés) en offrant une expérience utilisateur plus interactive.

9.3 Exploration de données

L'exploration de données désigne l'application d'algorithmes pour la découverte d'informations valides, nouvelles et utiles à partir de données. L'exploration de données a pris de l'importance à la fin des années 1990 et a été reconnue comme distincte des méthodes statistiques antérieures. Les statistiques traditionnelles se concentraient sur la collecte des données nécessaires et suffisantes pour répondre catégoriquement à une question spécifique. L'exploration de données était généralement appliquée à des données réadaptées afin de trouver des réponses approximatives ou des ajustements probabilistes à des formes. L'exploration de données est considérée comme l'étape de modélisation algorithmique du processus KDD complet. Fruit des premiers efforts en matière d'exploration de données, un consortium a pu détailler toutes les étapes de l'exploration de données dans la norme industrielle CRISP-DM publiée en 2000^[28]. L'exploration de données englobe un éventail de techniques, dont les arbres de décision, le regroupement et la classification. Avec l'émergence des technologies des mégadonnées au milieu des années 2000, l'application des algorithmes ne pouvait plus être séparée du stockage des données, et l'échantillonnage minutieux a cédé la place à un traitement plus rapide à forte consommation de données. Ces changements ont mené à une nouvelle description de la version mégadonnées du processus de cycle de vie de la KDD en tant qu'activités de science des données. Bien que la KDD et la découverte de connaissances soient des termes courants en IA, ce qu'un ordinateur produit ce ne sont pas des connaissances, mais des informations.

9.4 Planification

La planification est une sous-discipline de l'IA. Elle est essentielle aux applications industrielles et importantes dans de nombreux domaines d'activité, tels que le management du risque, les soins de santé, les robots industriels collaboratifs, la cybersécurité, les assistants cognitifs et la défense.

La planification permet à la machine de trouver automatiquement une séquence procédurale d'actions pour atteindre certains objectifs tout en optimisant certaines mesures de performance. Du point de vue de la planification, un système occupe un certain état. L'exécution d'une action peut modifier l'état du système et la séquence d'actions proposée par la planification peut rapprocher le système de l'état initial de l'état cible.

10 Applications des systèmes d'IA

10.1 Généralités

Étant donné que les systèmes d'IA sont capables d'aider ou, dans certains cas, d'automatiser la prise de décision, de fournir des recommandations et d'aider à l'automatisation de certaines tâches, ils ont des applications dans différents secteurs d'activité, y compris:

- agriculture et élevage;
- automobile;
- banque et technologie financière;
- défense;

- éducation;
- énergie;
- soins de santé;
- droit;
- industrie manufacturière;
- médias et divertissement;
- réalité mixte;
- secteur public;
- commerce de détail et marketing;
- sécurité;
- technologie spatiale;
- télécommunications.

Des exemples d'applications de l'IA sont fournis aux [Paragraphes 10.2 à 10.4](#).

10.2 Détection de la fraude

La fraude est l'utilisation de la tromperie dans un but lucratif. La fraude apparaît dans de nombreux domaines, y compris:

- contrefaçon d'argent et de documents;
- vol de cartes de crédit et de documents;
- communications personnelles telles que les courriers électroniques;
- falsification ou usurpation d'identité.

Ci-après des exemples d'applications de l'IA utilisées dans la détection de la fraude:

- identification des opérations par carte de crédit frauduleuses;
- identification des demandes de prêt ou de crédit frauduleuses;
- identification des déclarations frauduleuses de sinistre aux assurances;
- identification des accès aux comptes frauduleux.

10.3 Véhicules automatisés

Les véhicules sans conducteur sont d'après les projections appelés à se banaliser. Actuellement, de nombreuses technologies assistées par l'IA sont appliquées dans les automobiles en tant que fonctions d'aide à la conduite. Ci-après des exemples d'applications de l'IA dans les véhicules:

- optimisation de l'itinéraire (par exemple: recherche de l'itinéraire le plus rapide en fonction des conditions de circulation du moment);
- changement de voie automatisé;
- évitement d'objets (par exemple: actionnement automatique des freins, de l'accélérateur et de la direction sur la base de l'interprétation de signaux provenant de caméras et de capteurs de télédétection par laser);

- trajet entièrement automatisé d'un point A à un point B.

Les véhicules automatisés s'appuient sur des technologies d'IA telles que la vision artificielle et la planification.

10.4 Maintenance prédictive

Contrairement à la maintenance préventive, qui repose sur la durée de vie des composants (par exemple: le temps moyen entre défaillances), la maintenance prédictive prévoit l'entretien ou le remplacement des composants sur la base d'observations de leur comportement ou de leurs performances actuels et de leur durée de vie prévue. Ci-après des exemples d'applications de l'IA dans la maintenance prédictive:

- détection des interstices sous les voies ferrées (qui peuvent entraîner des déraillements);
- détection de l'asphalte fissuré ou endommagé;
- détection des roulements défaillants dans les moteurs électriques;
- détection des fluctuations anormales de la puissance dans les systèmes d'alimentation électrique.

Annexe A (informative)

Correspondance entre le cycle de vie d'un système d'IA et la définition du cycle de vie d'un système d'IA de l'OCDE

Les instruments juridiques de l'OCDE ont publié une «Recommandation du Conseil sur l'intelligence artificielle»^[26].

Ce qui suit est inclus dans le présent document:

«Sur proposition du Comité de la politique de l'économie numérique:

I. CONVIENT qu'aux fins de la présente Recommandation, les termes techniques suivants s'entendent comme suit:

- cycle de vie d'un système d'IA: Le cycle de vie d'un système d'IA comporte plusieurs phases:
 - i) la phase de «conception, données et modèles», séquence dépendante du contexte comprenant la planification et la conception, la collecte et le traitement des données, ainsi que la construction du modèle;
 - ii) la phase de «vérification et validation»;
 - iii) la phase de «déploiement»;
 - iv) la phase «d'exploitation et de suivi».

Ces phases présentent souvent un caractère itératif et ne suivent pas nécessairement un ordre séquentiel. La décision de mettre un terme à l'utilisation d'un système d'IA peut intervenir à n'importe quel stade de la phase d'exploitation et de suivi».

et

«Robustesse, sûreté et sécurité:

- a) les systèmes d'IA devraient être robustes, sûrs et sécurisés tout au long de leur cycle de vie, de sorte que, dans des conditions d'utilisation normales ou prévisibles, ou en cas d'utilisation abusive ou de conditions défavorables, ils soient à même de fonctionner convenablement, et ne fassent pas peser un risque de sécurité démesuré;
- b) pour ce faire, les acteurs de l'IA devraient veiller à la traçabilité, notamment pour ce qui est des ensembles de données, des processus et des décisions prises au cours du cycle de vie des systèmes d'IA, afin de permettre l'analyse des résultats produits par lesdits systèmes d'IA et le traitement des demandes d'information, compte tenu du contexte et de l'état de l'art de la technologie;
- c) les acteurs de l'IA devraient, selon leurs rôles respectifs, le contexte et leur capacité à agir, appliquer de manière continue une approche systématique de la gestion du risque, à chaque phase du cycle de vie des systèmes d'IA, afin de gérer les risques y afférents, notamment ceux liés au respect de la vie privée, à la sécurité numérique, à la sûreté et aux biais».

La [Figure A.1](#) montre comment cette définition du cycle de vie d'un système d'IA peut être mise en correspondance avec le cycle de vie d'un système d'IA décrit à [l'Article 6](#).

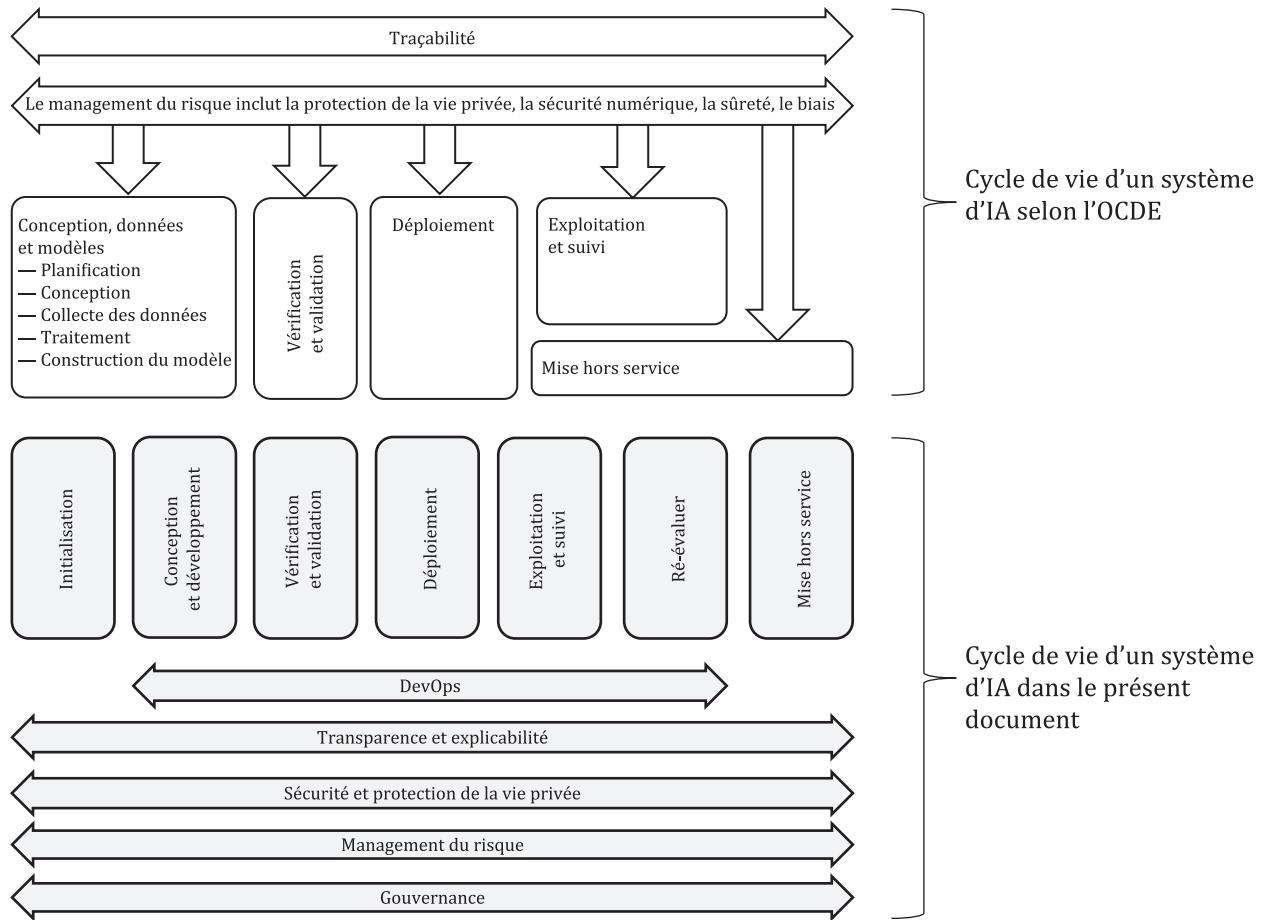


Figure A.1 — Correspondance avec le cycle de vie d'un système d'IA de l'OCDE

Bibliographie

- [1] ISO/IEC 12207:2008, *Ingénierie des systèmes et du logiciel — Processus du cycle de vie du logiciel*
- [2] ISO/IEC IEEE 15288:2015, *Ingénierie des systèmes et du logiciel — Processus du cycle de vie du système*
- [3] ISO/IEC IEEE 15289:2019, *Ingénierie des systèmes et du logiciel — Contenu des articles d'information du cycle de vie (documentation)*
- [4] ISO/IEC 17788:2014, *Technologies de l'information — Informatique en nuage — Vue d'ensemble et vocabulaire*
- [5] ISO/IEC 17789:2014, *Technologies de l'information — Informatique en nuage — Architecture de référence*
- [6] ISO/IEC 20546:2019, *Technologies de l'information — Mégadonnées — Vue d'ensemble et vocabulaire*
- [7] ISO/IEC 20547-3:2020, *Technologies de l'information — Architecture de référence des mégadonnées — Partie 3: Architecture de référence*
- [8] ISO/IEC 20889:2018, *Terminologie et classification des techniques de dé-identification de données pour la protection de la vie privée*
- [9] ISO/IEC 20924:2021, *Technologies de l'information — Internet des objets (IdO) — Vocabulaire*
- [10] ISO/IEC 23053, *Cadre méthodologique pour les systèmes d'intelligence artificielle (IA) utilisant l'apprentissage machine*
- [11] ISO/IEC/TR 23188:2020, *Information technology — Cloud computing — Edge computing landscape*
- [12] ISO/IEC 23894, *Technologies de l'information — Intelligence artificielle — Recommandations relatives au management du risque*
- [13] ISO/IEC/TR 24027:2021, *Technologie de l'information — Intelligence artificielle (IA) — Tendance dans les systèmes de l'IA et dans la prise de décision assistée par l'IA*
- [14] ISO/IEC/TR 24028:2020, *Technologies de l'information — Intelligence artificielle — Examen d'ensemble de la fiabilité en matière d'intelligence artificielle*
- [15] ISO/IEC/TR 24029-1:2021, *Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*
- [16] ISO/IEC 27040:2015, *Technologie de l'information — Techniques de sécurité — Sécurité de stockage*
- [17] ISO/IEC 30141:2018, *Architecture de référence de l'Internet des objets (IoT RA)*
- [18] GRAVES A., Abdel-rahman Mohamed, Geoffrey E. Hinton, Speech recognition with deep recurrent neural networks, IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, DOI:10.1109/ICASSP.2013.6638947.
- [19] JENA A., Reasoners and rule engines: Jena inference support, <https://jena.apache.org/documentation/inference/index.html>
- [20] Artificial Intelligence Methodologies and Their Application to Diabetes. <https://pubmed.ncbi.nlm.nih.gov/28539087/>
- [21] ELMAN Jeffrey L.“, Finding structure in time.” Cognitive science **14**.2 (1990): 179-211

- [22] HOCHREITER Sepp, SCHMIDHUBER Juergen“, Long short-term memory.” Neural computation **9**.8 (1997): 1735-1780
- [23] JAPANESE SOCIETY OF ARTIFICIAL INTELLIGENCE. AI Map Beta, https://www.ai-gakkai.or.jp/pdf/aimap/AIMap_EN_20190606.pdf
- [24] ZADEH L.A.“, Soft computing and fuzzy logic,” IEEE Software, 1994, vol. **11**, issue 6
- [25] RIGLA M., GEMA GARCÍA-SÁEZ B., Pons, M., Artificial Intelligence Methodologies and Their Application to Diabetes Hernando, Journal of diabetes science and technology, 2018, DOI:10.1177/1932296817710475.
- [26] Recommandation du Conseil sur l'intelligence artificielle. <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449>
- [27] ROZENBLIT J.W., Cognitive computing: Principles, architectures, and applications. In: Proc. 19th European Conf. on Modelling and Simulation (ECMS) (2005)
- [28] S. C., The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); **5**:13 — 22
- [29] Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach (3rd Edition) (Essex, England: Pearson, 2009)
- [30] TAXONOMY AND DEFINITIONS FOR TERMS RELATED TO DRIVING AUTOMATION SYSTEMS FOR ON-ROAD MOTOR VEHICLES. SAE — On-Road Automated Driving (ORAD) committee, https://saemobilus.sae.org/content/J3016_201806/
- [31] ISO/IEC IEEE 24765:2017, *Ingénierie des systèmes et du logiciel — Vocabulaire*
- [32] ISO/IEC 2382:2015, *Technologies de l'information — Vocabulaire*
- [33] ISO 16439:2014, *Information et documentation — Méthodes et procédures pour évaluer l'impact des bibliothèques*
- [34] ISO/IEC 2382-28:1995, *Technologies de l'information — Vocabulaire — Partie 28: Intelligence artificielle — Notions fondamentales et systèmes experts*
- [35] ISO 8373:2012, *Robots et composants robotiques — Vocabulaire*
- [36] ISO 20252:2019, *Études de marché, études sociales et d'opinion, y compris insights et analytique de données — Vocabulaire et exigences de service*
- [37] ISO/IEC 29100:2011/Amd1:2018, *Technologies de l'information — Techniques de sécurité — Cadre privé — Amendement 1: Clarifications*
- [38] ISO/IEC 38500:2015, *Technologies de l'information — Gouvernance des technologies de l'information pour l'entreprise*
- [39] ISO/IEC 27000:2018, *Technologies de l'information — Techniques de sécurité — Systèmes de management de la sécurité de l'information — Vue d'ensemble et vocabulaire*
- [40] ISO 31000:2018, *Management du risque — Lignes directrices*
- [41] ISO/IEC 27042:2015, *Technologies de l'information — Techniques de sécurité — Lignes directrices pour l'analyse et l'interprétation des preuves numériques*
- [42] ISO 17100:2015, *Services de traduction — Exigences relatives aux services de traduction*
- [43] ISO/IEC 15944-8:2012, *Technologies de l'information — Vue opérationnelle d'affaires — Partie 8: Identification des exigences de protection de la vie privée en tant que contraintes externes sur les transactions d'affaires*

- [44] ISO 5127:2017, *Information et documentation — Fondations et vocabulaire*
- [45] ISO/IEC 20071-11, *Technologies de l'information — Accessibilité du composant interface utilisateur — Partie 11: Lignes directrices pour les textes alternatifs pour images*

