



Outils de mise en œuvre de l'intelligence artificielle au gouvernement du Canada

Luc Gagnon

Sous-ministre adjoint principal et dirigeant principal de la technologie
Services partagés Canada / Gouvernement du Canada



Services partagés Shared Services
Canada Canada

Canada

Sujets

- L'aventure de la démocratisation de l'intelligence artificielle (IA)
- Enjeux
- Outils de mise en œuvre de l'IA au gouvernement du Canada (GC)
- Outils de mise en œuvre d'une infrastructure d'AI susceptible d'intéresser Services partagés Canada (SSC)

Gouvernement du Canada et IA

Cours sur l'IA donnés en 1994 et en 1997 aux scientifiques du GC par Geoffrey Hinton, lauréat du prix Turing de 2018

ARTIFICIAL NEURAL NETWORKS

Ottawa, Sept 8 & 9 1994

Geoffrey Hinton
Department of Computer Science
University of Toronto

Copyright © Geoffrey Hinton, 1994.
Permission is hereby granted to the Government of Canada
to use this material for internal use only.

UNSUPERVISED NEURAL NETWORKS

Ottawa, Oct 3 1997

Geoffrey Hinton
Department of Computer Science
University of Toronto

Copyright © Geoffrey Hinton, 1997.
Permission is hereby given for reproduction
for internal use by the Government of Canada.

Démocratisation de l'intelligence artificielle

Une aventure guidée par les percées de la recherche sur les algorithmes et soutenue par la loi de Moore...

Voilà un quart de siècle environ, les perceptrons multicouches ne comptaient que de trois à dix couches

- Tout était écrit en langages procéduraux, accompagnés de quelques bibliothèques
- Seuls les chercheurs spécialisés en IA se servaient de la technologie
- Peu de données disponibles aux fins de formation et d'essai
- Extrême rareté des ordinateurs spécialisés adaptés à l'IA

Et aujourd'hui, en 2019

- Divers cadres de l'IA, comme Tensorflow et Theano, rendent les réseaux neuronaux profonds (RNP) et l'apprentissage machine (AM) disponibles sans codage des fonctions d'IA, et le nombre de couches est passé à plusieurs centaines
- Infrastructure d'infonuagique
 - Des unités de traitement tensoriel et graphique (TPU et GPU) sont disponibles dans le nuage, quelle que soit la plateforme utilisée
 - On parle maintenant de bassins de données
- Plateformes et services d'IA
 - IBM Watson, AM Amazon Web Service (AWS), Google Cloud AutoML, modèles personnalisables comme You Only Look Once (YOLO), AWS SageMaker, Microsoft Machine Learning, etc.
 - Services sans AM requis : traitement du langage naturel, vision par ordinateur, traduction, prévisions, mégadonnées

Pourquoi tarde-t-on à adopter l'IA?

1. Manque de compétences en science des données et en IA
2. Quantités insuffisantes de données étiquetées destinées à l'AM
3. Besoin de plateformes et d'outils d'IA largement disponibles, adaptables et tenus à jour
4. Souveraineté des données et gouvernance de la gestion de la sécurité des applications d'IA
5. Sélection des problèmes appropriés à résoudre avec l'IA
6. Ensembles de données et bancs d'essai significatifs pour confirmer un rendement réel

Outils de mise en œuvre de l'AI au GC



Shared Services
Canada

Services partagés
Canada

Canada

Infrastructure et plateformes d'IA fournies par le programme « Le nuage d'abord »

- Programme d'infonuagique complet en cours au GC



Services d'IA hébergés

- Couvrent le traitement du langage naturel, la vision par ordinateur, les prévisions, la traduction, la compréhension du discours oral
- Certains tirent profit de modèles préformés et se perfectionnent à l'aide d'ensembles de données de moindre envergure fournis par les utilisateurs
- Orientent l'étiquetage efficace des données pour un usage ultérieur
- Conçus à l'intention des développeurs de logiciels peu familiers avec les algorithmes
- Peuvent être fournis par le marché des services d'infonuagique du GC
- Exploitent souvent des interfaces de programmation d'applications en conformité avec la Stratégie numérique du GC
- Exemples d'utilisation :
 - Interfaces vocales conversationnelles destinées aux applications du gouvernement
 - Services d'agents automatiques (bots)
 - Reconnaissance faciale
- Ces outils de mise en œuvre de services d'IA hébergés répondent aux enjeux 1, 2, 4 et 5*

* Enjeux

1. Manque de compétences en science des données et en IA
2. Quantités insuffisantes de données étiquetées destinées à l'AM
3. Besoin de plateformes et d'outils d'IA largement disponibles, adaptables et tenus à jour
4. Souveraineté des données et gouvernance de la gestion de la sécurité des applications d'IA
5. Sélection des problèmes appropriés à résoudre avec l'IA
6. Ensembles de données et bancs d'essai significatifs pour confirmer un rendement réel

Cadres d'IA avancés

- Fondés sur des cadres en libre accès, comme Theano, scikit-learn, Tensorflow, CNTK et MXNET
- Surtout utilisés par les scientifiques des données et les spécialistes de l'IA du gouvernement du Canada
- Facilitent la collaboration en recherche et en développement
- Réduisent le temps consacré au codage et permettent aux chercheurs d'expérimenter sans délai divers algorithmes et différentes architectures de RNP
- Fournissent souvent des mises en œuvre déjà construites qui permettent de réaliser des bancs d'essai expérimentaux largement disponibles
- Exemples d'utilisation :
 - Laboratoire d'IA destiné aux scientifiques dans les ministères à vocation scientifique
 - Analyse avancée des données en science des données
- Cet outil de mise en œuvre des cadres de l'IA répond aux enjeux 3 et 6*

* Enjeux

1. Manque de compétences en science des données et en IA
2. Quantités insuffisantes de données étiquetées destinées à l'AM
3. Besoin de plateformes et d'outils d'IA largement disponibles, adaptables et tenus à jour
4. Souveraineté des données et gouvernance de la gestion de la sécurité des applications d'IA
5. Sélection des problèmes appropriés à résoudre avec l'IA
6. Ensembles de données et bancs d'essai significatifs pour confirmer un rendement réel

Stratégie du GC en matière de données reposant sur l'IA

- La stratégie en matière de données des ministères doit reposer sur des besoins à court et long termes :
 - L'apprentissage machine a besoin d'un fort volume de données étiquetées
 - Au moins 5 000 exemples étiquetés par catégorie pour atteindre un niveau de rendement acceptable
 - Au moins 10 millions d'exemples étiquetés pour atteindre le niveau de rendement humain
 - Une stratégie conviviale en matière de données passe par une stratégie en matière de métadonnées
 - Manque d'uniformité des normes d'étiquetage des données
 - Envisager l'utilisation de données synthétiques pour créer des capacités d'IA
 - plus économique, temps de collecte plus court
 - Si le modèle de synthèse est complet, les données peuvent être beaucoup plus riches (polyvalentes et spécialisés) que les données recueillies dans le monde réel
 - Étiquetage automatique – précision sans faille
 - Peuvent remplacer les données de nature délicate impossibles à obtenir
- Cet outil de mise en œuvre de la stratégie du GC en matière de données reposant sur l'IA répond aux enjeux 2, 4 et 6*

* Enjeux

1. Manque de compétences en science des données et en IA
2. Quantités insuffisantes de données étiquetées destinées à l'AM
3. Besoin de plateformes et d'outils d'IA largement disponibles, adaptables et tenus à jour
4. Souveraineté des données et gouvernance de la gestion de la sécurité des applications d'IA
5. Sélection des problèmes appropriés à résoudre avec l'IA
6. Ensembles de données et bancs d'essai significatifs pour confirmer un rendement réel

Outils de mise en œuvre de l'infrastructure d'IA



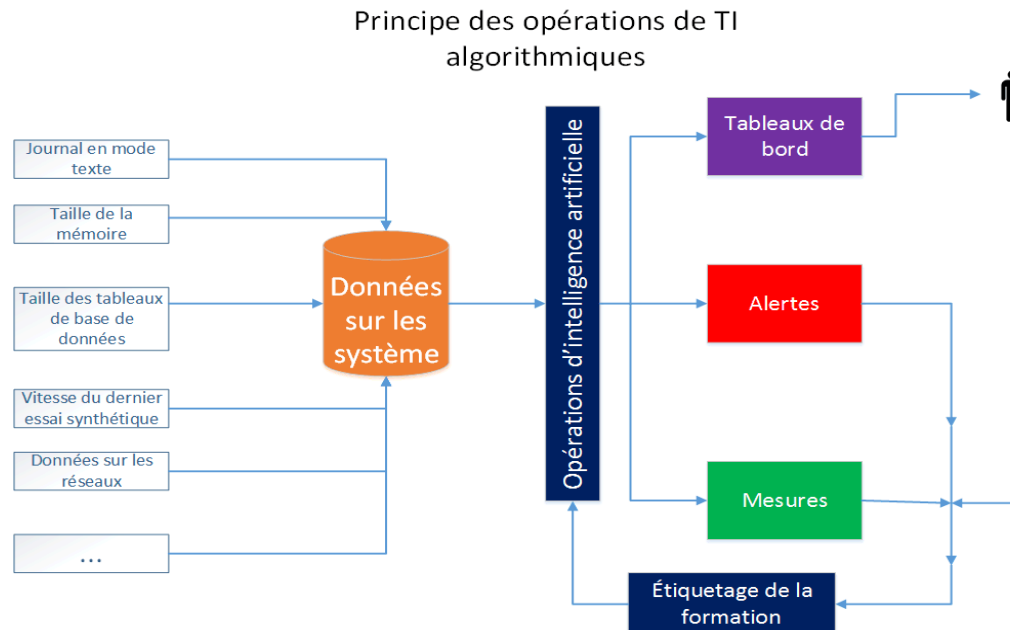
Shared Services
Canada

Services partagés
Canada

Canada

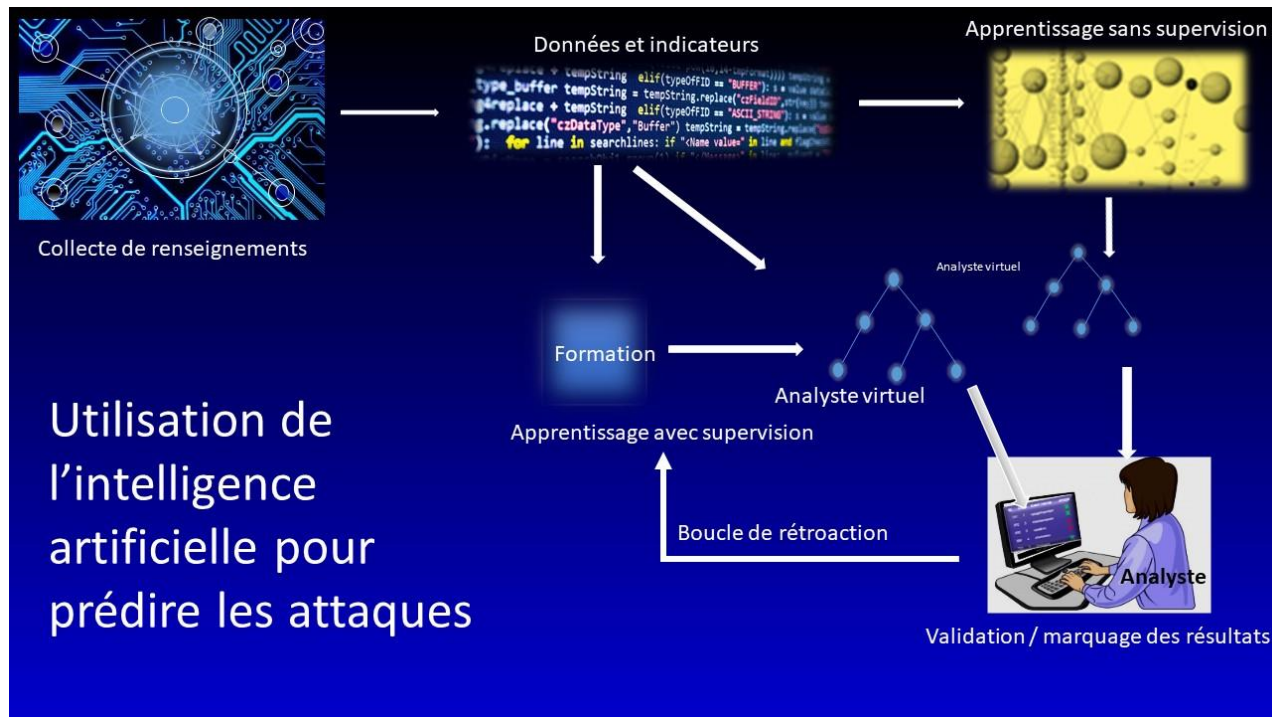
Fonctionnement des réseaux et des centres de données avec l'IA : opérations de TI algorithmiques (Op IT Alg)

- Les gens les plus doués ne sont plus capables de composer avec la quantité d'événements que génère leur environnement informatique; en revanche, les Op IT Alg peuvent en traiter des millions et permettent ainsi aux humains de se concentrer sur les choses importantes
- Ajout de capacités diagnostiques à la surveillance avancée, comme la surveillance du rendement des applications (SRA) et la surveillance et le diagnostic du rendement des réseaux (SDRR)
- Exemple typique de visualisation et de diagnostic par l'apprentissage machine
- Les plateformes d'OP IT Alg améliorent le fonctionnement des technologies de l'information par l'apprentissage machine et la visualisation
- Diagnostic posé plus rapidement et prévention proactive des problèmes techniques, en corrélant différentes sources de données
- Diagnostic en temps réel facilitant le travail des équipes de débogage et réduisant le temps moyen de réparation (TMR)
- Peuvent servir à automatiser les mesures nécessaires pour corriger un problème



IA et cybersécurité

- Seule l'automatisation intelligente peut contrer la vitesse de propagation et le degré de perfectionnement des cybermenaces en y réagissant dans des délais significatifs à l'échelle informatique (c'est-à-dire mesurés en secondes et non en jours, semaines ou mois)
- Investir dans les bassins de données contenant de vastes quantités de données pertinentes à analyser
- Corrélation en temps réel d'un large éventail d'événements
- Mise en file d'attente des incidents et établissement de leur niveau de priorité
- Orchestration et automatisation de la sécurité et intervention (OASI)
- Analyse du comportement des utilisateurs et des entités (ACUE)
- Analyse prédictive des risques et des interventions fondée sur l'analyse des menaces et de la posture de défense





Des questions?