



# Conception de trousse d'outils de traitement du langage naturel au sein de l'industrie : Leçons apprises et messages à retenir concernant les assistants virtuels

6 décembre 2021

Gias Uddin, Analyse avancée, BSIF

*Le contenu de cette présentation est fondé sur mon expérience de travail et mes opinions.  
Il ne reflète pas forcément le point de vue de mon employeur, le BSIF.*



BSIF  
OSFI

Canada

Je mène des recherches et je crée des applications ayant une incidence dans le monde réel à partir d'outils et de techniques propres à la science des données.

Études



**Queen's**  
UNIVERSITY

**Maîtrise en  
informatique  
2006 – 2008**

Modèle mathématique de confiance pour les logiciels axés sur le service



**McGill**  
UNIVERSITY

**Doctorat en informatique  
2010 – 2018**

Production automatique de résumés des sentiments et d'examens de données provenant des médias sociaux



**UNIVERSITY OF  
CALGARY**

**Professeur adjoint (poste menant à la permanence)  
Génie électrique et génie logiciel  
2020 à aujourd'hui**

Administrateur fondateur du laboratoire DISA (Data Intensive Software Analytics) ayant comme tâche de concevoir un logiciel d'apprentissage automatique responsable



**recognia**<sup>®</sup>

a TRADING CENTRAL company

**Ingénieur de données  
2008 – 2010**

Développement de logiciels d'analyse des investissements faisant appel aux cours des actions et aux bilans des sociétés



**Chercheur et ingénieur logiciel**

**IBM Watson Analytics  
2011 – 2015**

Déterminer automatiquement les intentions dans les demandes textuelles des utilisateurs, et produire des résultats en vue de fournir une réponse à ces demandes au moyen du traitement du langage naturel (TLN)



**Scientifique des données  
2016 – 2018**

Traitement de millions d'étiquettes de courrier numérisées de Postes Canada; et production d'analyses (détection des fraudes)



**BANK OF CANADA  
BANQUE DU CANADA**

**Scientifique des données  
principal**

**Bureau des données et des  
statistiques, 2018 – 2020**

Modèle de gouvernance des données et de prévision pour les microdonnées de crédit à la consommation de TransUnion



**OSFI  
BSIF**

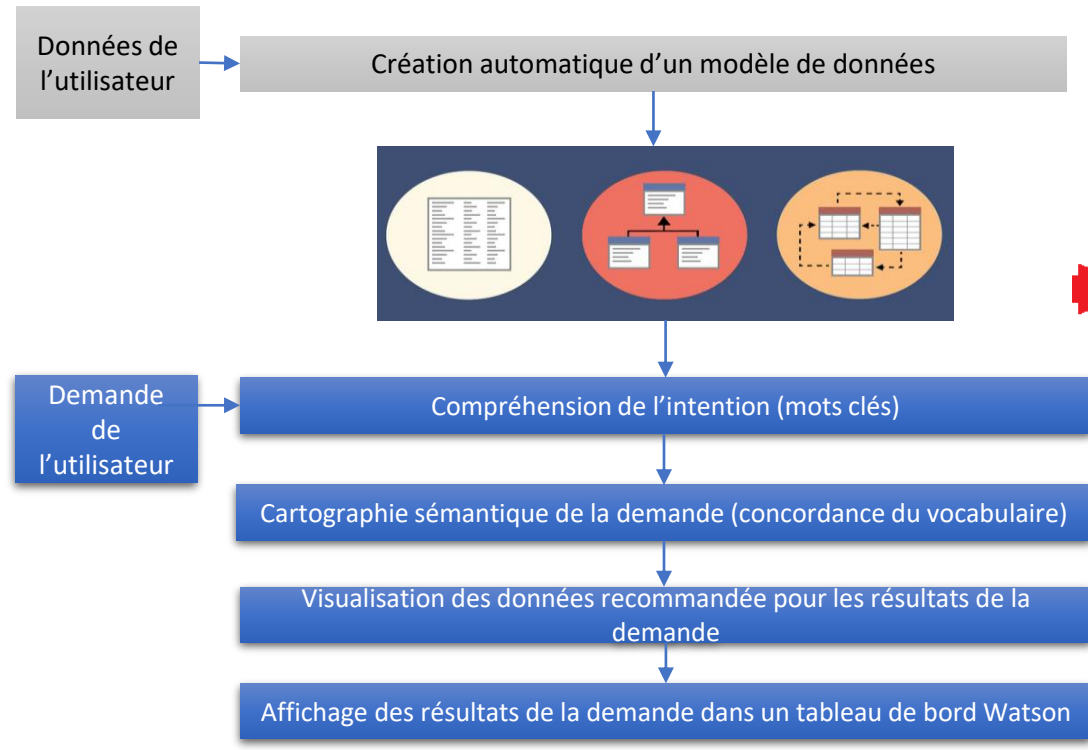
**Gestionnaire**

**Analyse avancée  
2021 – aujourd'hui**

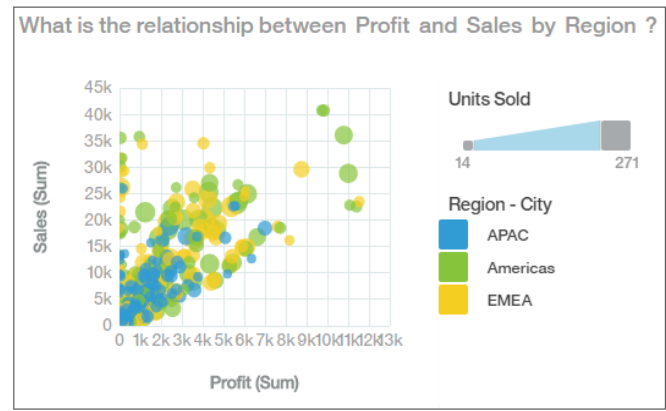
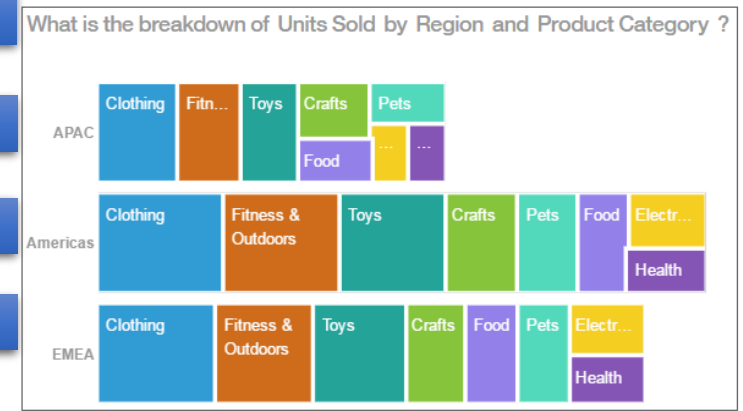
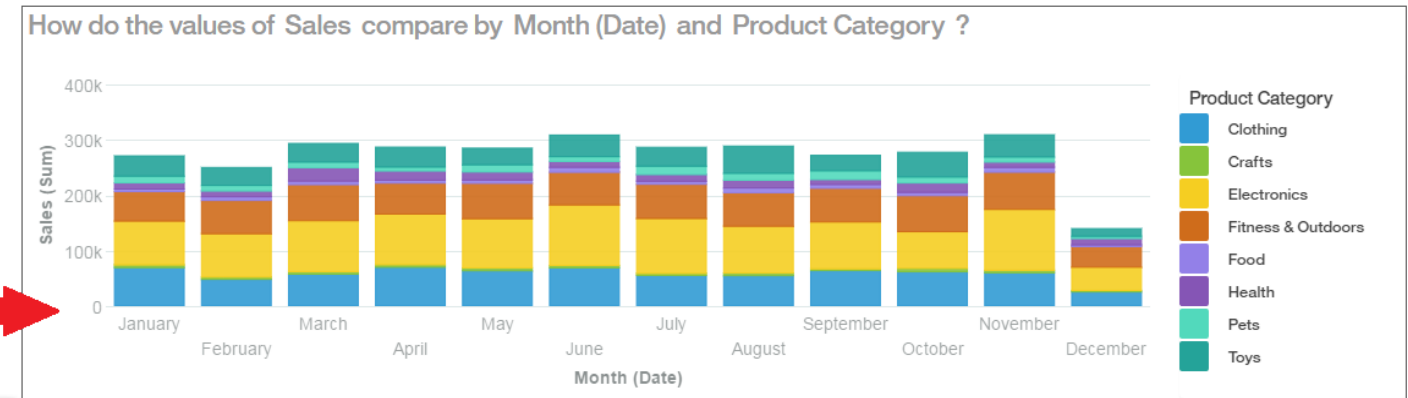
Conception et développement d'outils technologiques de supervision à l'aide de techniques faisant appel à l'IA, à l'apprentissage automatique et au TLN

Industrie et  
gouvernement

# Compréhension automatique des « intentions » des utilisateurs à partir de demandes en langage naturel, et production d'une réponse



## Product Sales Performance



L'équipe de conception de produits était composée de centaines de réalisateurs de logiciels, d'architectes et de gestionnaires de produits de plusieurs pays sur deux continents (Amérique du Nord et Europe)

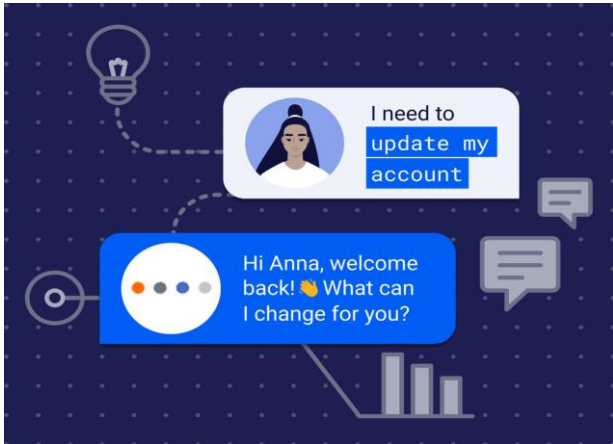
Il a fallu plus de deux ans de travaux de planification et de conception pour mettre au point la première version utilisable par les consommateurs

Une fois opérationnel, a été utilisé par des millions et des millions d'utilisateurs

L'équipe principale responsable du traitement du langage naturel était basée à Ottawa (la détection de l'intention était considérée comme étant une réalisation clé chez IBM)

A eu une incidence sur la conception de multiples produits dérivés au sein d'IBM (p. ex. analyse des perspectives des consommateurs d'IBM, Orchestrate d'IBM)

# Leçons apprises et messages à retenir concernant la création d'assistants virtuels



Outil

Il est presque impossible de trouver un outil mis au point par une tierce partie et qui est en mesure d'accomplir toutes les tâches de TLN permettant de donner suite à des objectifs opérationnels précis

Il faut souvent apporter d'importantes adaptations aux outils existants pour qu'ils puissent donner suite à des besoins précis

Technique

Dans toute analyse textuelle, le modèle d'apprentissage automatique (AA) n'est qu'un élément entrant dans de nombreuses tâches qui sont axées d'abord, non pas sur l'AA à proprement parler, mais sur le TLN/les données

La tâche de détection de l'intention ou toute autre tâche faisant intervenir le TLN (p. ex. classification) peut être un problème propre au domaine et nécessiter des modèles spécialisés

Opérationnalisation

L'opérationnalisation des modèles d'AA requiert souvent un changement holistique en raison de la nécessité d'intégrer ces modèles ainsi que les flux/pipelines opérationnels dans les systèmes logiciels traditionnels

L'adoption des outils de TLN doit s'inscrire dans une démarche axée sur la personne, en mettant l'accent sur les besoins particuliers à combler et en actualisant continuellement les outils en fonction de l'évolution des exigences

IA responsable

Il peut y avoir un changement fondamental d'approche entre la manière dont un modèle est conçu à l'étape de l'exploration et la manière dont il sera utilisé dans un cadre réel

Si un assistant virtuel intègre des données inadéquates ou repose sur des techniques ou des outils inappropriés, les conséquences négatives ne se limiteront pas forcément à la production de résultats inadéquats



**L'assistant virtuel Tay de Microsoft a été désactivé 16 heures après son lancement, car des trolls actifs sur Twitter lui avaient appris à utiliser de l'argot et à tenir des propos inappropriés.**

[https://fr.wikipedia.org/wiki/Tay\\_\(intelligence\\_artificielle\)](https://fr.wikipedia.org/wiki/Tay_(intelligence_artificielle))