

**100% ACTUAIRES &
100% DATA SCIENCE**

INSTITUT DES
ACTUAIRES



Comment améliorer l'expérience client et la maîtrise des risques avec la data science ?

Des expérimentations à l'industrialisation

29 / NOV / 2019

Hôtel Marriott Rive Gauche
Paris 14ème

Nicolas MARESCAUX

Marc RAYMOND

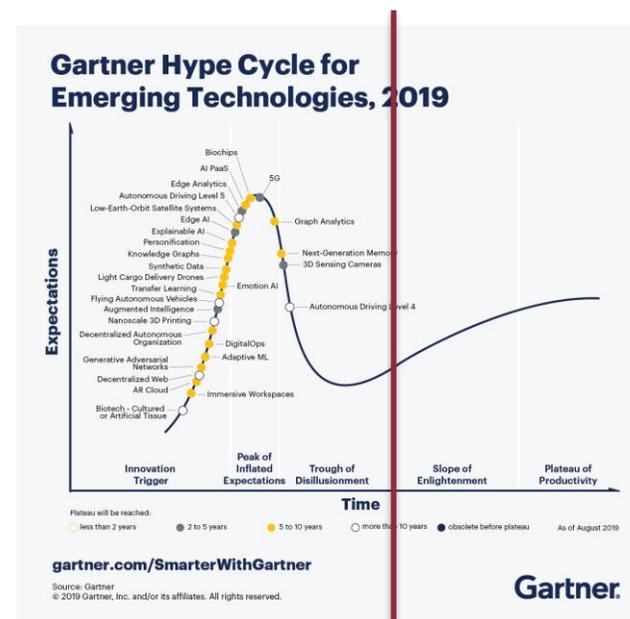
La problématique et les principaux domaines d'application

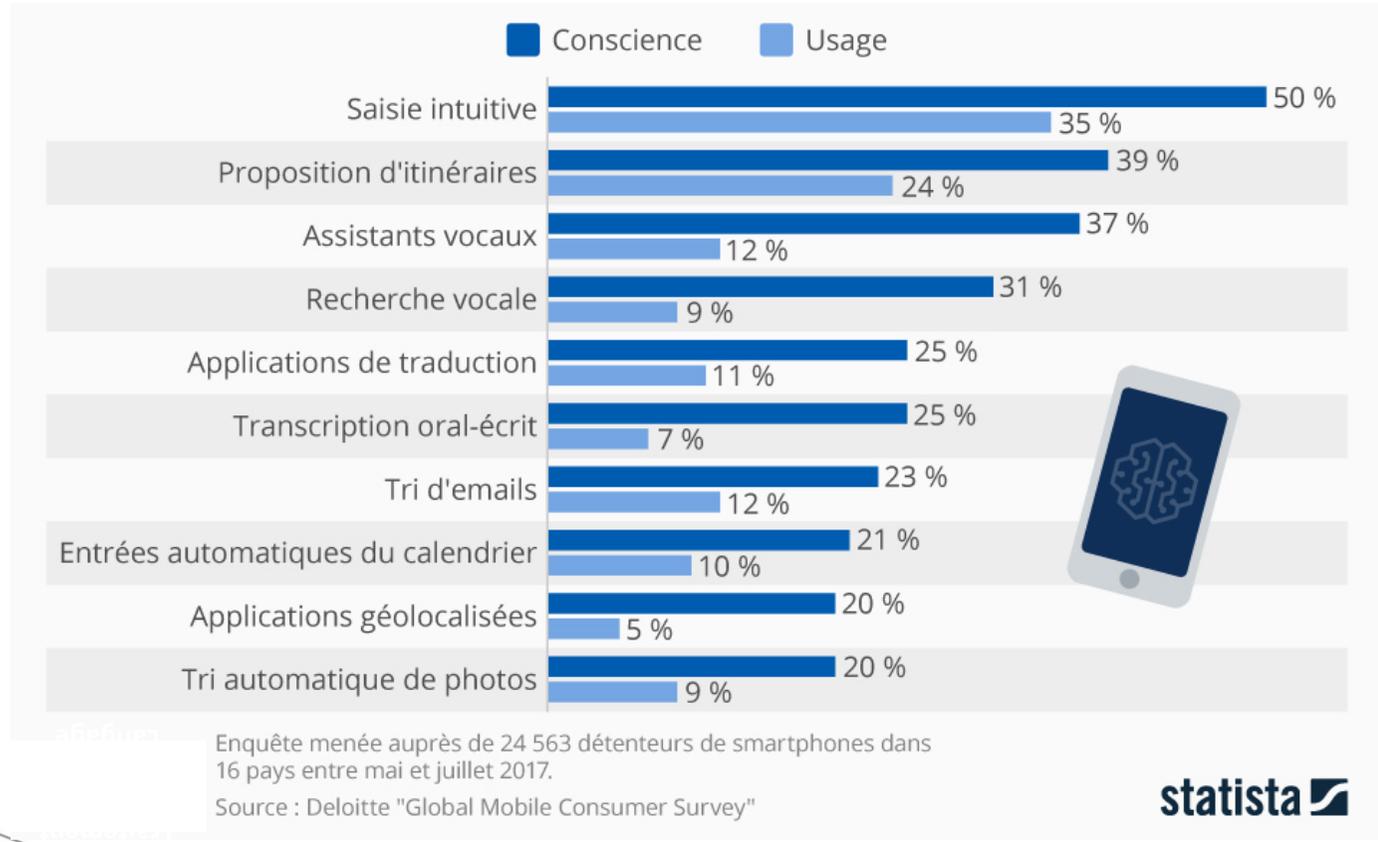


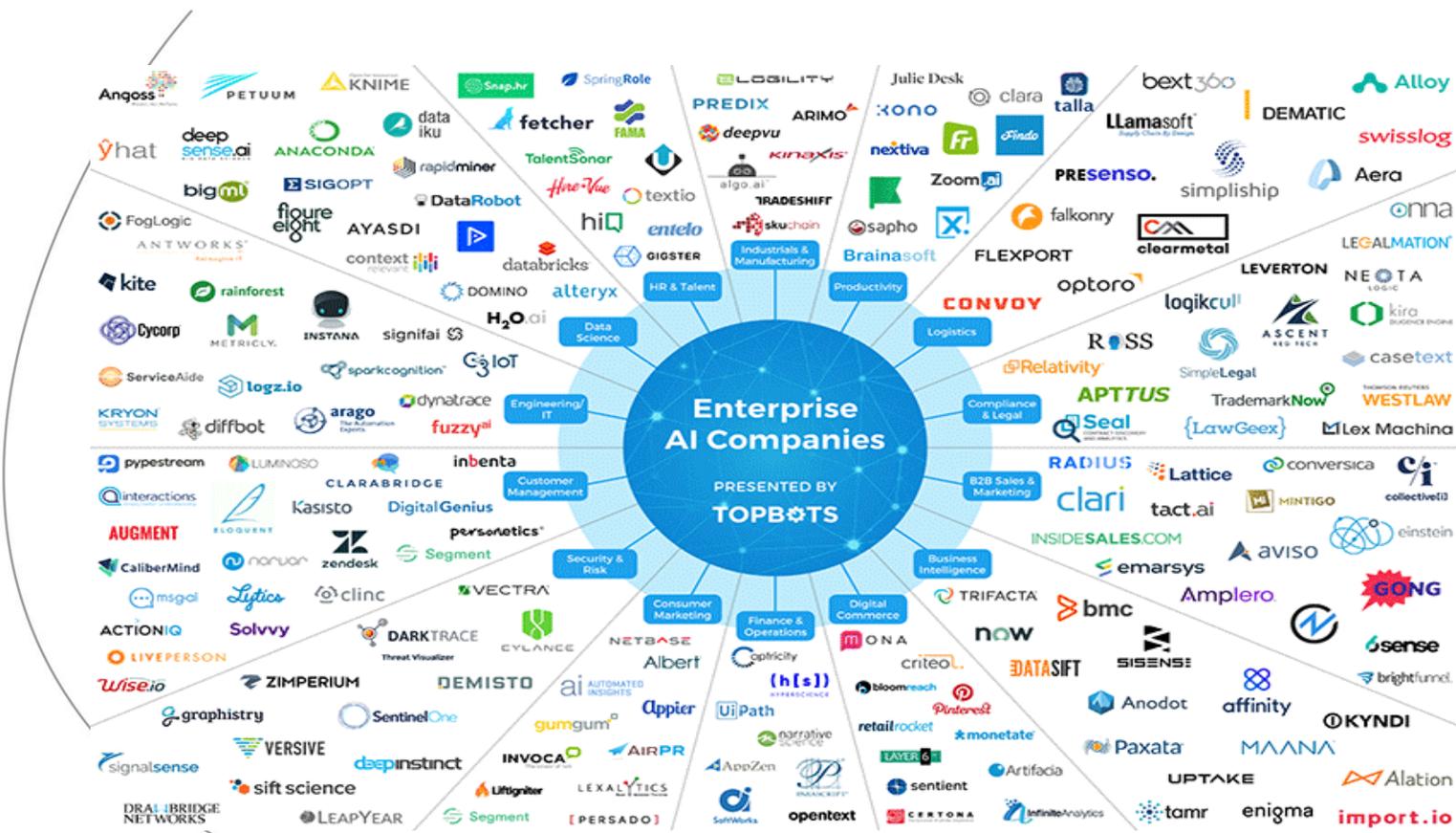
© marketoonist.com

- Des technologies expérimentées depuis une petite dizaine d'années
- Parfois spécifiques, mais souvent capitalisant sur les recherches dans d'autres secteurs de l'économie
- Rendues possibles par l'émergence de nouvelles technologies, de nouvelles données et de l'augmentation de la puissance de calcul des ordinateurs
- Qui après de nombreux « POC » entrent dans une phase de mise en œuvre opérationnelle
- Dans certains processus, mais pas tous
- Des processus en lien avec le client mais aussi internes
- Avec encore de nombreuses pistes à explorer
- Et systématiquement la nécessité de mettre en place des démarches d'accompagnement du changement

29/11/2019







- 278 M€ de levée de fonds en 2017
- 1,5 Md€ d'investissement par le gouvernement français sur la période 2018–2022
- Croissance trimestrielle moyenne de l'ordre de 80 % entre T2 2016 et T2 2017
- 18 masters dispensant des formations IA
- 5300 chercheurs



- 90 milliards de dollars pour le marché mondial de l'IA d'ici 2025
- 1730 jeunes pousses consacrées à l'IA dans 70 pays avec une levée de fonds moyenne de 7,5 millions de dollars par entreprise



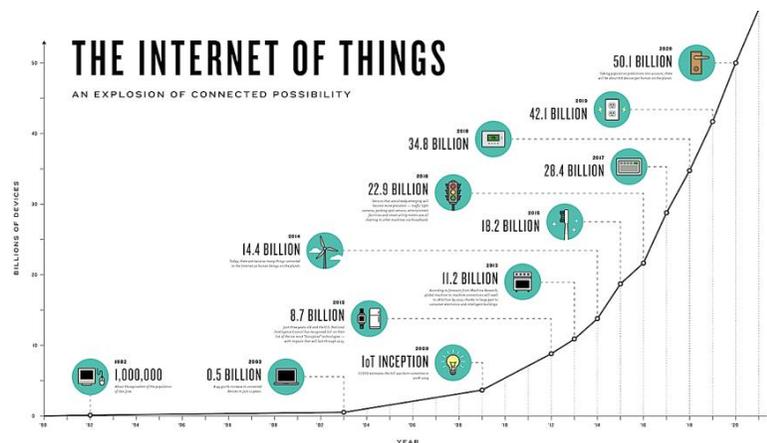
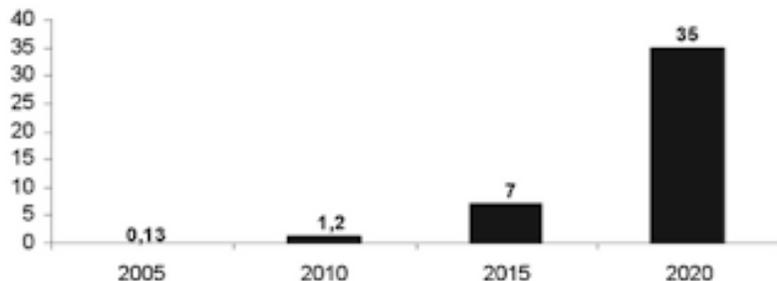
Source : livre blanc finance innovation

DOMAINES		THÉMATIQUES IMPACTÉES PAR LA DATA SCIENCE	TECHNOLOGIE	Niveau de maturité
AXE DÉVELOPPEMENT	AVANT VENTE	<ul style="list-style-type: none"> Détermination de populations cibles et d' offres personnalisées Meilleure connaissance des clients et optimisation globale du parcours client Ciblage des campagnes marketing 	ML ML ML	ChatBot assuré
	SOUSCRIPTION	<ul style="list-style-type: none"> Aide en temps réel Optimisation du processus de souscription Optimisation du taux de concrétisation des devis Traitement automatisé des pièces nécessaires à la souscription 	CB ML- LN - RN ML RI	Automatisation des processus
	VALEUR CLIENT	<ul style="list-style-type: none"> Analyse de l'élasticité au prix d'achat, intégration dans le « competitive pricing » Analyse du risque de résiliation Anticipation des évolutions de comportement 	ML ML ML	Gestion des sinistres
AXE TECHNIQUE	PRODUITS ET SERVICES	<ul style="list-style-type: none"> Conception de nouveaux services et de nouvelles offres Intégration des nouvelles technologies dans les offres actuelle Anticipation des évolutions du risque 	ML LN - RI - LNL ML	Automatisation de la souscription
	TARIFICATION, SUIVI TECHNIQUE	<ul style="list-style-type: none"> Modification du tarif par l'intégration de « scorings » multi-critères Modification du tarif par l'intégration de données externes nouvelles Identification des segments déficitaires et modulation des majorations tarifaires 	ML - BD ML - BD ML - BD	ChatBot vente
	GESTION DES CONTRATS	<ul style="list-style-type: none"> Réponse automatique aux questions des assurés Personnalisation de la relation client KYC - LCBFT Traitement automatisé de documents 	CB ML- LN - RN ML CB RI RI - RN	Utilisation des données IOT
AXE GESTION	SINISTRES	<ul style="list-style-type: none"> Estimation du degré d'urgence et priorisation des dossiers à traiter Assistance à la détermination du coût des sinistres Identification des comportements atypiques, 	LN RN - RI ML	Fraude
				Personnalisation de l'offre

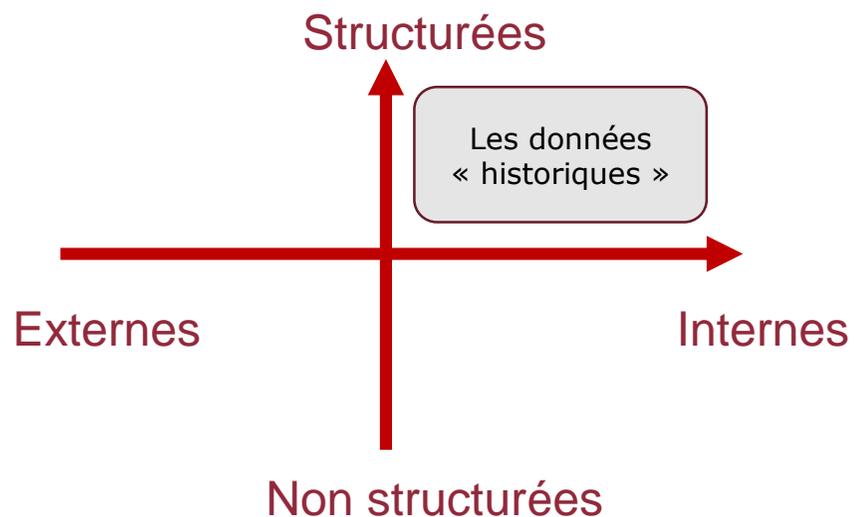
ML : machine learning CB : chatbot LN : langage naturel RI : reconnaissance d'image RN : réseaux de neurones

Les technologies sous-jacentes

Volume mondial de données (zettaoctets)



Les données

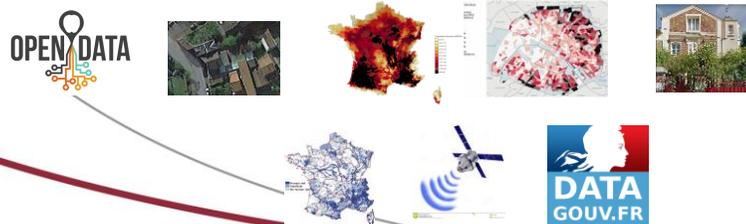


Une réglementation de plus en plus contraignante

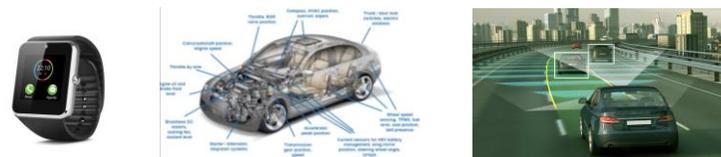
Le web Scraping

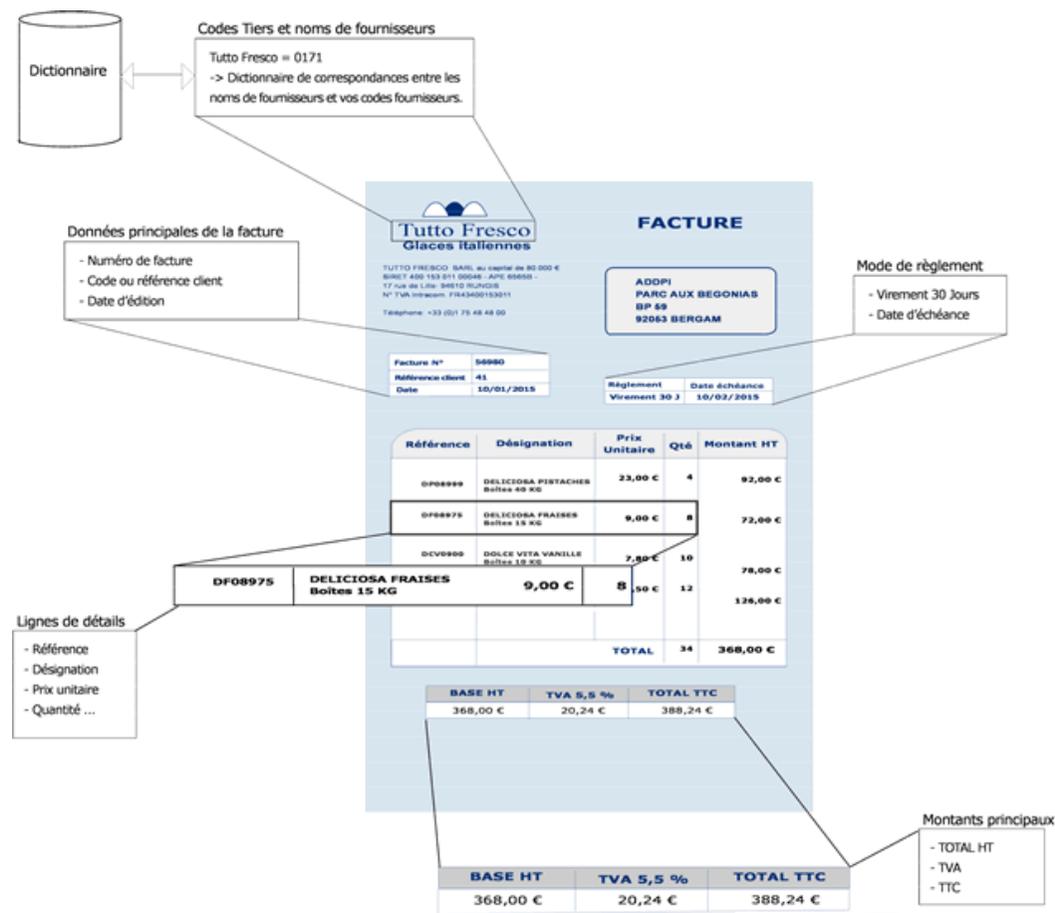
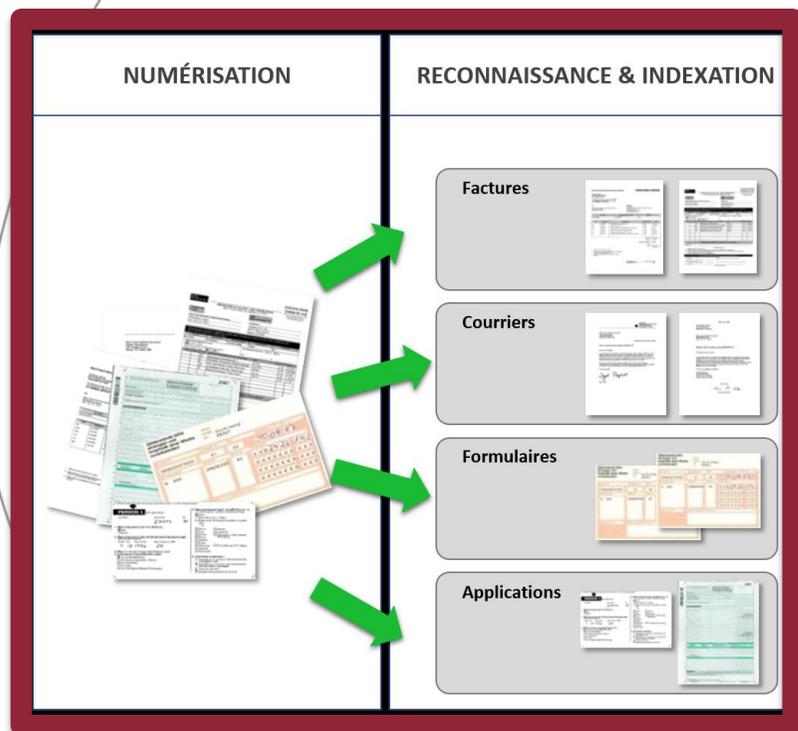


Les bases de données externes

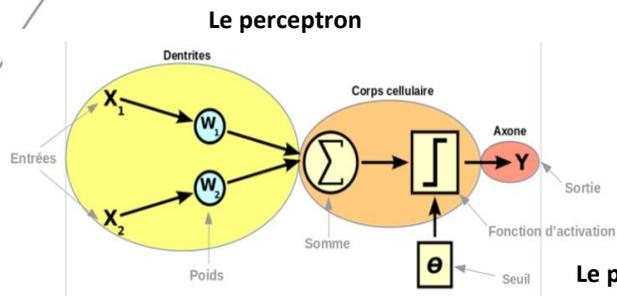


Les données embarquées

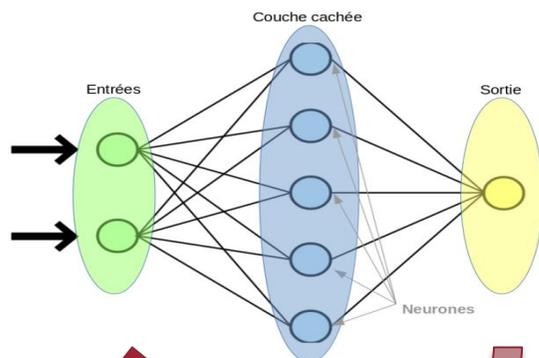




Le perceptron est inventé en 1957 par F. Rosenblatt. Le perceptron est le réseau de neurone le plus simple, il est composé de neurones qui fonctionnent légèrement différemment que le neurone formel. Ces neurones sont appelés des unités linéaires à seuil (LTU, Linear Threshold Unit). Les LTU ont en entrées des nombres quelconques (contrairement aux neurones formels qui ont des entrées binaires) et chaque entrée est pondérée par un poids (w) :

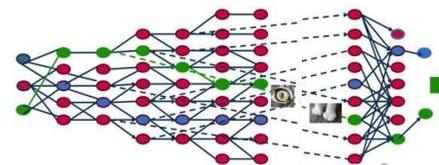


Le perceptron multicouche

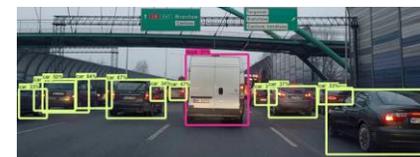
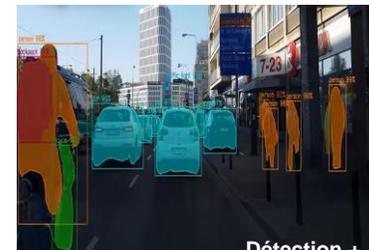
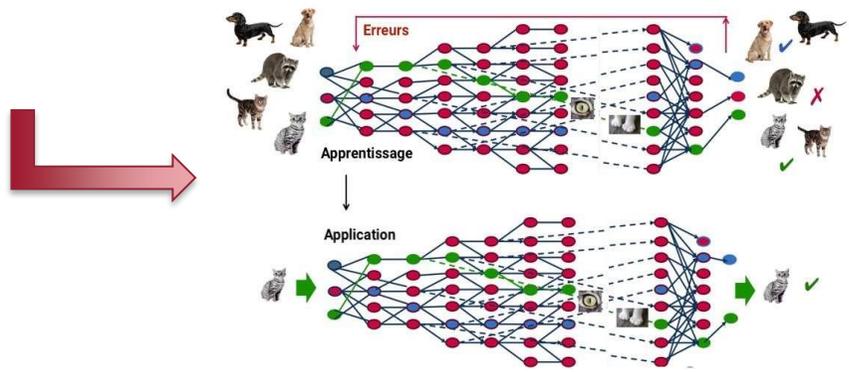
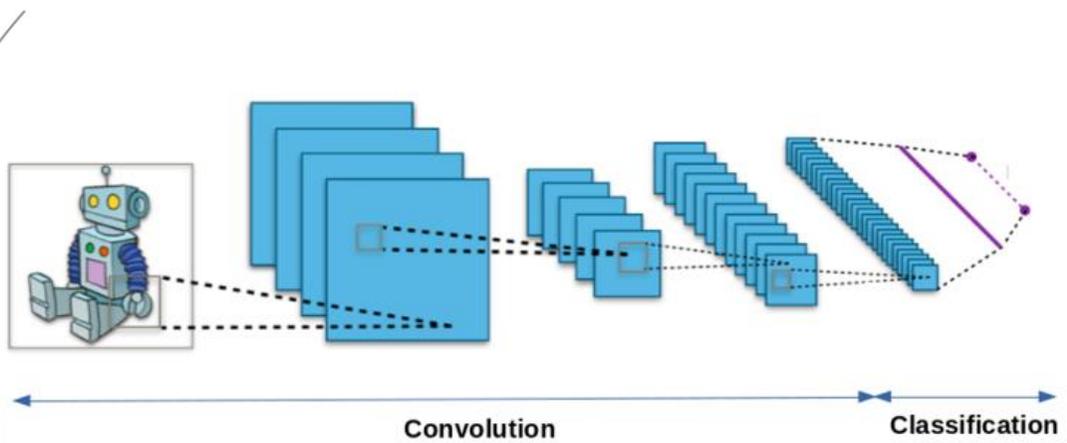


Mécanisme d'apprentissage

Rétropopagation de l'erreur
pour ajuster les poids
On itère jusqu'à ce que les erreurs
deviennent minimales

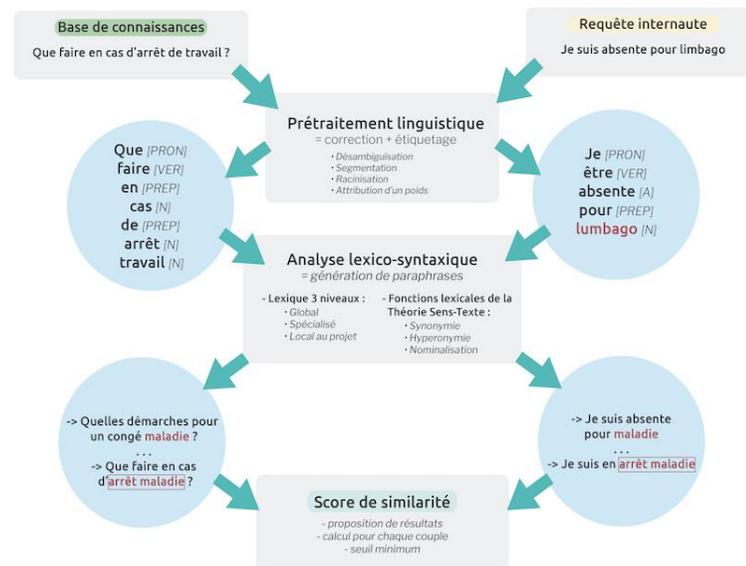
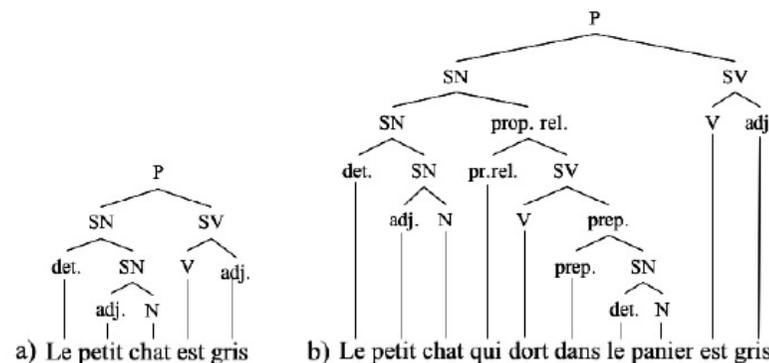


Chemin optimal



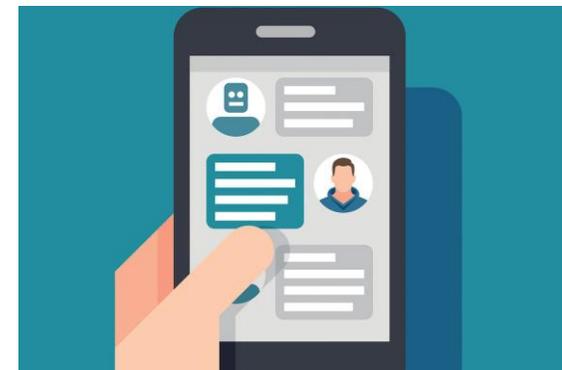
Dans le traitement du langage naturel, différents domaines d'application de la linguistique sont pris en compte :

- La **morphologie**, qui s'intéresse à la composition des mots et leur corrélation avec d'autres mots.
- La **syntaxe**, qui définit comment les mots sont agencés dans une phrase.
- La **sémantique**, qui correspond à la signification des mots et des groupes de mots
- La **pragmatique**, grâce à laquelle le contexte est pris en compte.



La démarche

1. Identifier et définir quelles seront les problématiques que les utilisateurs chercheront à résoudre
2. Rédiger une liste de questions que l'outil sera en mesure de comprendre et de gérer
3. Définir les réponses à ces questions
4. Imaginer la personnalité de l'assistant virtuel : son ton et son style sont définis en fonction du public auquel il s'adressera.
5. Développer techniquement l'outil



1. Reconnaître les mots et le sens de la phrase

2. Déterminer le sens selon le contexte

3. Prendre la bonne décision

Après avoir déterminé le sens de la question, le chatbot doit trouver la réponse à fournir. Pour cela, il peut faire appel à un arbre de décision, des bases de données, des conversations passées, etc.

4. Formuler la réponse

Enfin, le chatbot doit formuler sa réponse. Pour cela, il peut s'appuyer sur une bibliothèque de réponses prédéfinies ou la Génération de Langage Naturel.

Traitement
Automatique
du Langage

Arbres de
décision

Donnée



Information / compréhension



Réponse

Taxonomie des modèles de machine learning

Les modèles supervisés

Les modèles non supervisés

Linéaires

Non linéaires

K means

EM

Régressions

Perceptron

Combinés

Bagging

Boosting

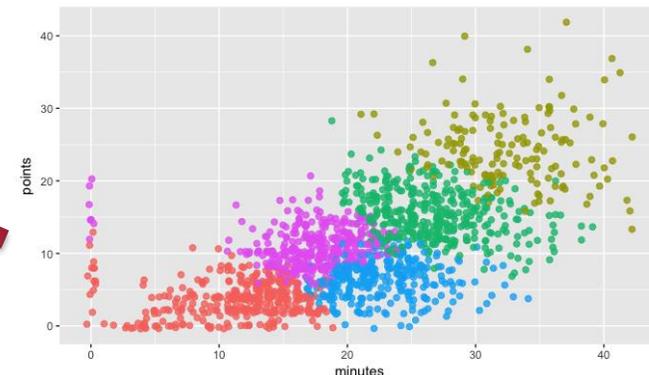
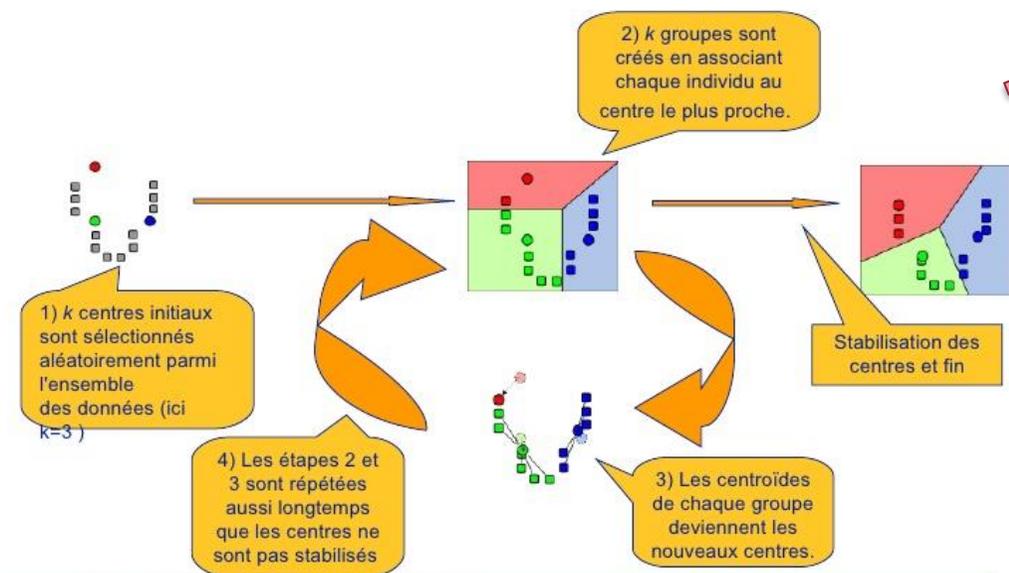
Random forest

Unitaires

CART

SVM

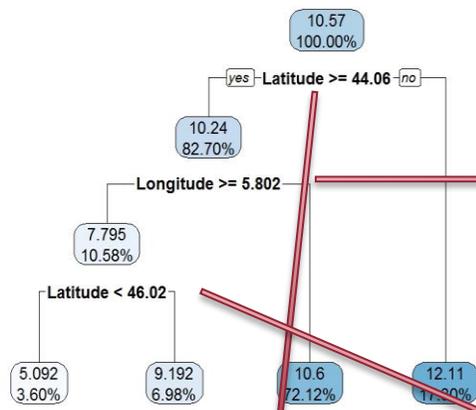
Les Kmeans



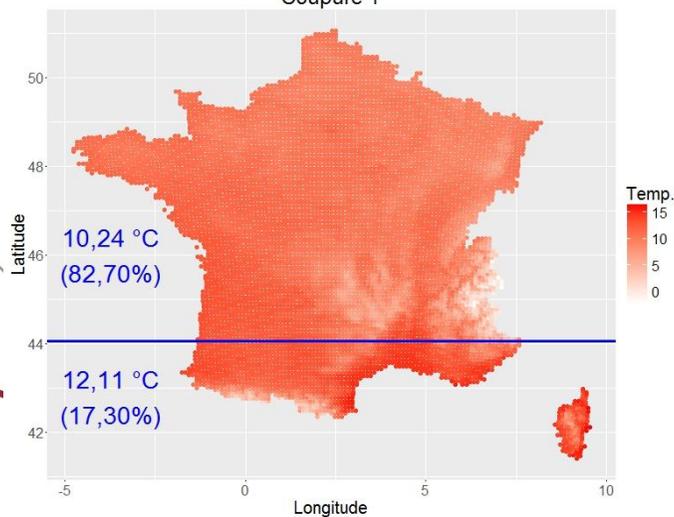
Ne travaille que sur des données continues

Suppose que les échelles aient été « normalisées »

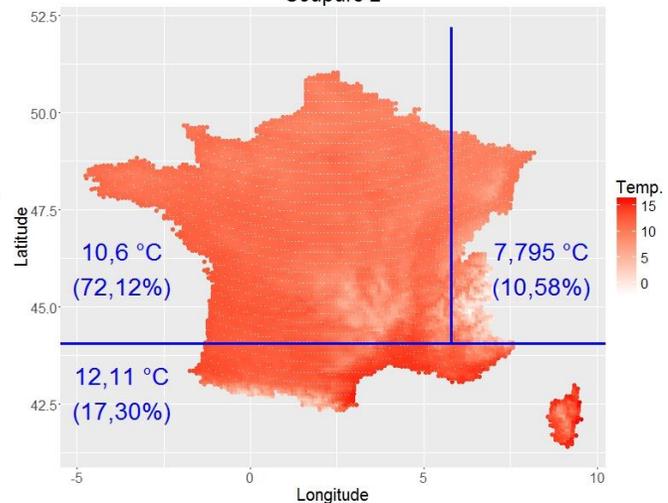
Les CARTs



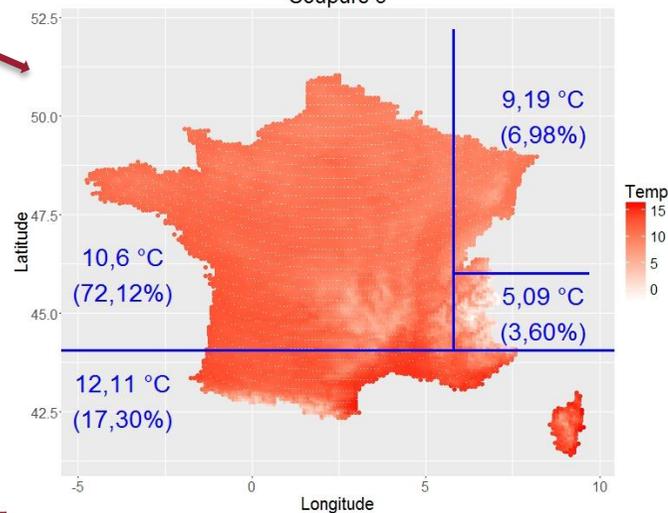
Coupure 1



Coupure 2

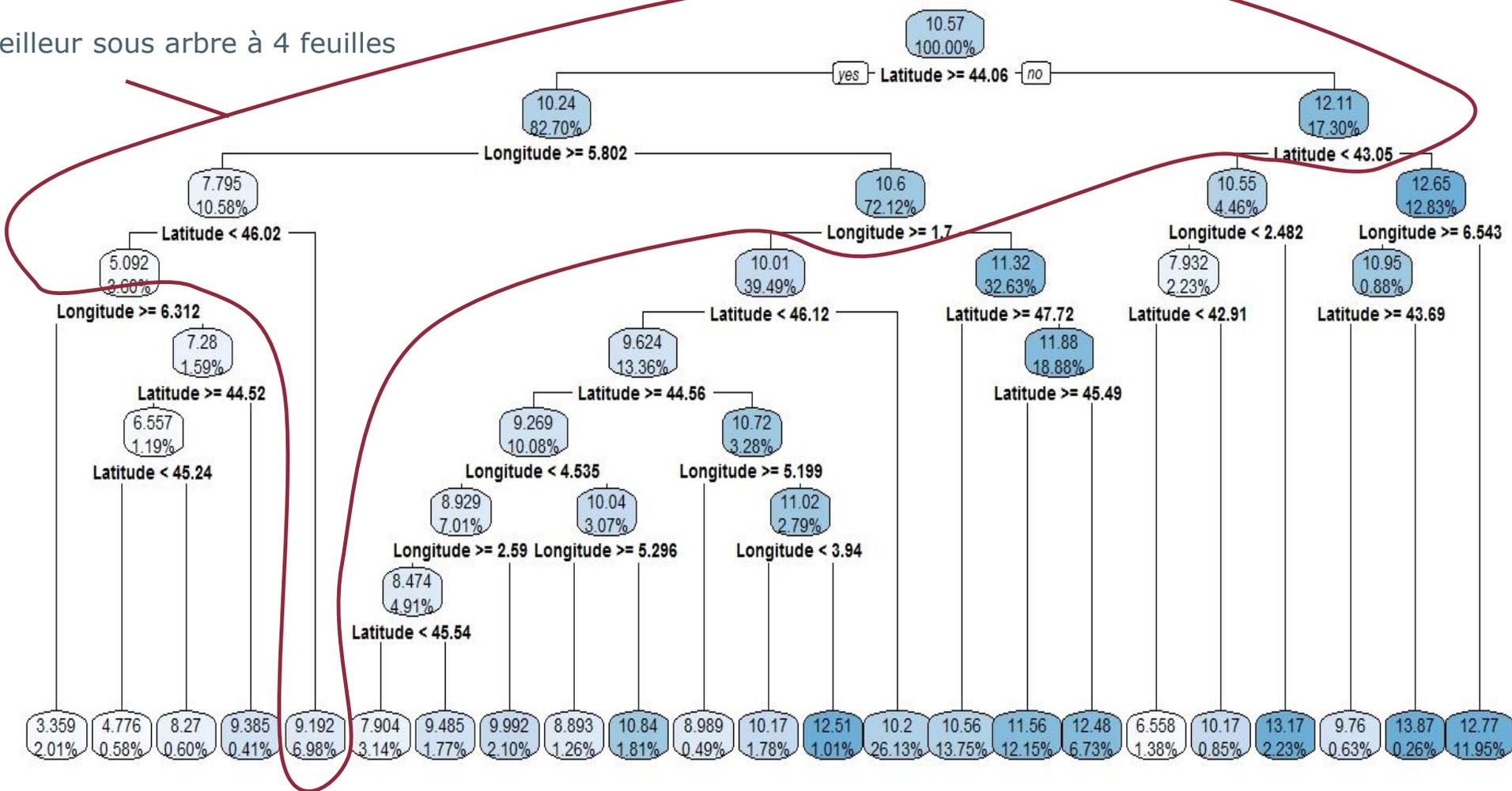


Coupure 3



Meilleur sous arbre à 23 feuilles

Meilleur sous arbre à 4 feuilles



Random forest

Succession de CARTs avec :

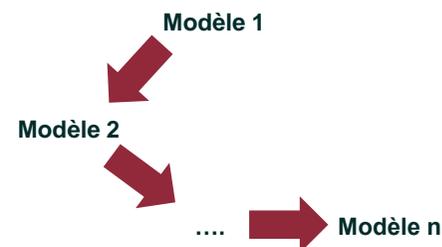
- sélection aléatoire de variables
- Sélection aléatoire d'individus



Agrégation des arbres $\hat{f}(\cdot, E_n) = \frac{1}{B} \sum_{k=1}^B \hat{f}(\cdot, E_n^k)$

Gradient Boosting

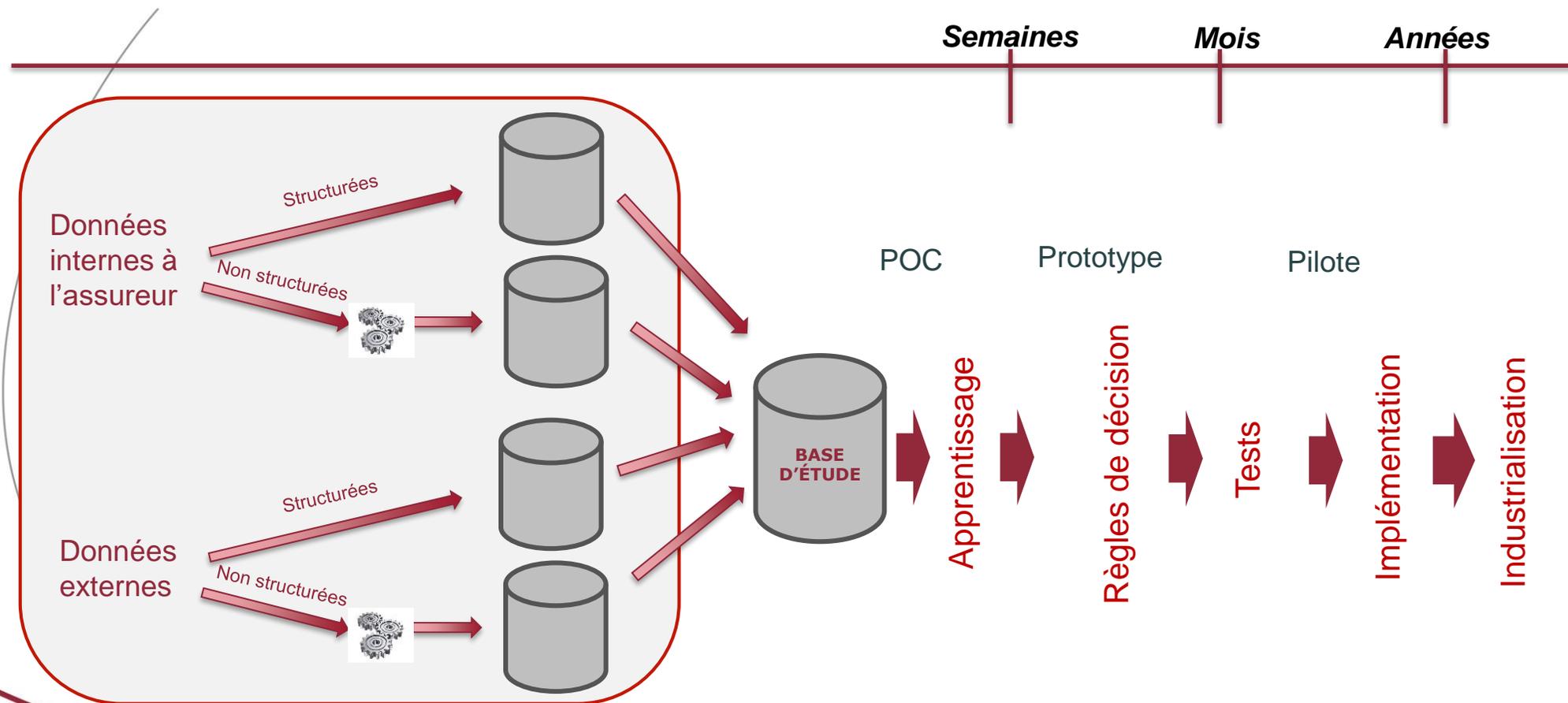
Succession de modèles sur les « résidus » des modèles précédents afin d'obtenir la meilleure explicabilité possible.



Un bon pouvoir prédictif mais une grande difficulté à interpréter les résultats

Des risques de sur interprétation en cas de nombre de données insuffisant

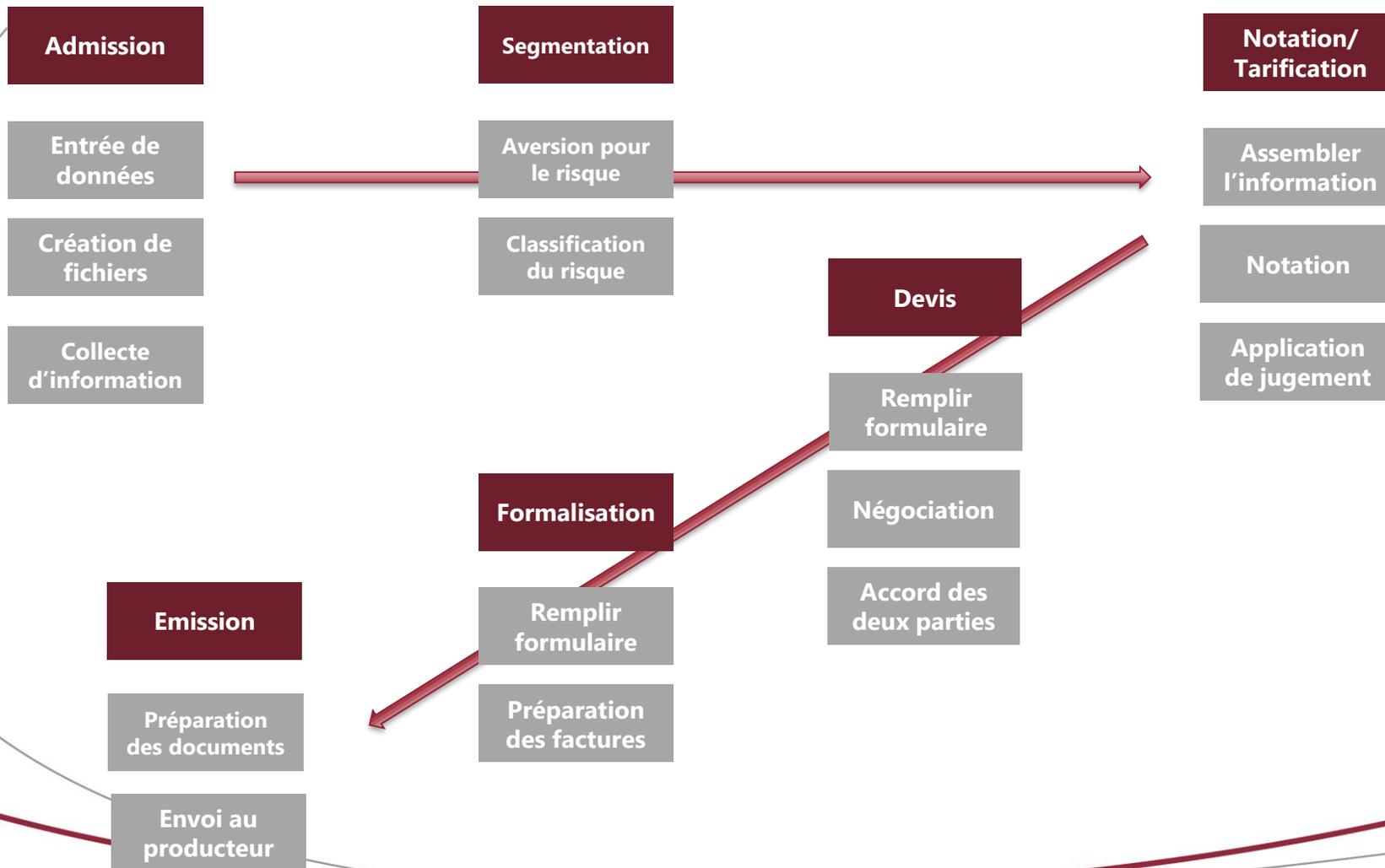
Complémentaire avec les modèles linéaires

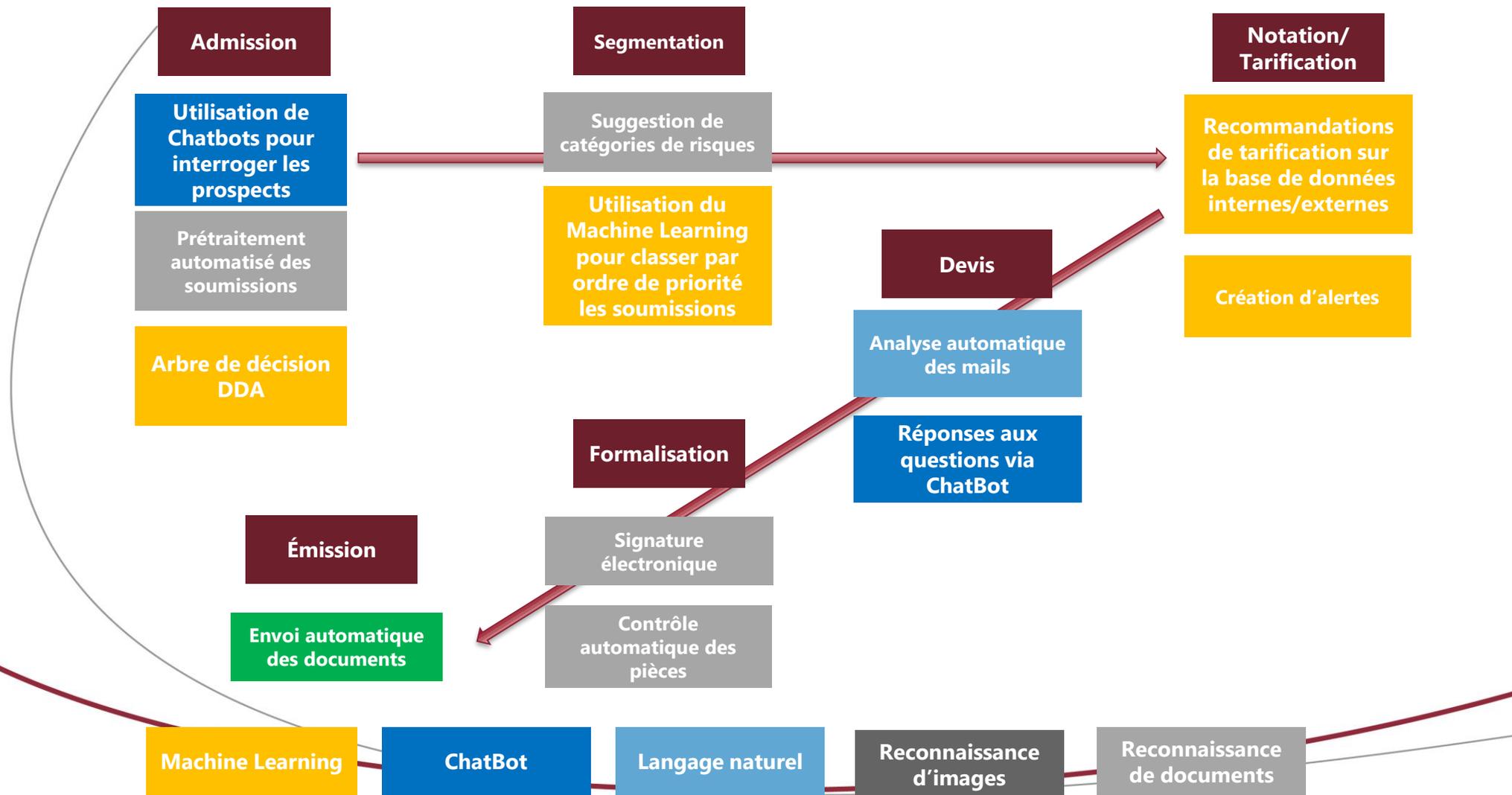


Apports sur les processus de l'assureur

	DOMAINES	THÉMATIQUES IMPACTÉES PAR LA DATA SCIENCE	TECHNOLOGIE
AXE DÉVELOPPEMENT	AVANT VENTE	<ul style="list-style-type: none"> Détermination de populations cibles et d' offres personnalisées Meilleure connaissance des clients et optimisation globale du parcours client Ciblage des campagnes marketing 	ML ML ML
	SOUSCRIPTION	<ul style="list-style-type: none"> Aide en temps réel Optimisation du processus de souscription Optimisation du taux de concrétisation des devis Traitement automatisé des pièces nécessaires à la souscription 	CB ML- LN - RN ML RI
	VALEUR CLIENT	<ul style="list-style-type: none"> Analyse de l'élasticité au prix d'achat, intégration dans le « competitive pricing » Analyse du risque de résiliation Anticipation des évolutions de comportement 	ML ML ML
AXE TECHNIQUE	PRODUITS ET SERVICES	<ul style="list-style-type: none"> Conception de nouveaux services et de nouvelles offres Intégration des nouvelles technologies dans les offres actuelle Anticipation des évolutions du risque 	ML LN - RI - LNL ML
	TARIFICATION, SUIVI TECHNIQUE	<ul style="list-style-type: none"> Modification du tarif par l'intégration de « scorings » multi-critères Modification du tarif par l'intégration de données externes nouvelles Identification des segments déficitaires et modulation des majorations tarifaires 	ML - BD ML - BD ML - BD
AXE GESTION	GESTION DES CONTRATS	<ul style="list-style-type: none"> Réponse automatique aux questions des assurés Personnalisation de la relation client KYC - LCBFT Traitement automatisé de documents 	CB ML- LN - RN ML CB RI RI - RN
	SINISTRES	<ul style="list-style-type: none"> Estimation du degré d'urgence et priorisation des dossiers à traiter Assistance à la détermination du coût des sinistres Identification des comportements atypiques, 	LN RN - RI ML

La souscription





Instantanéité

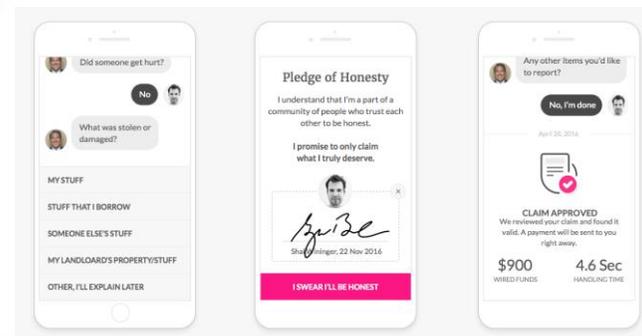
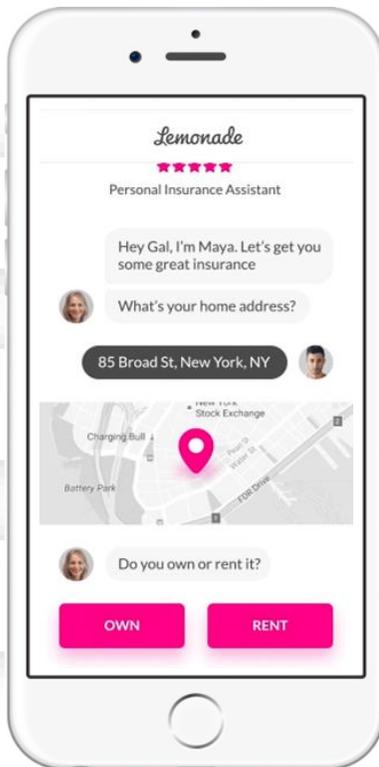
À la demande

À l'usage

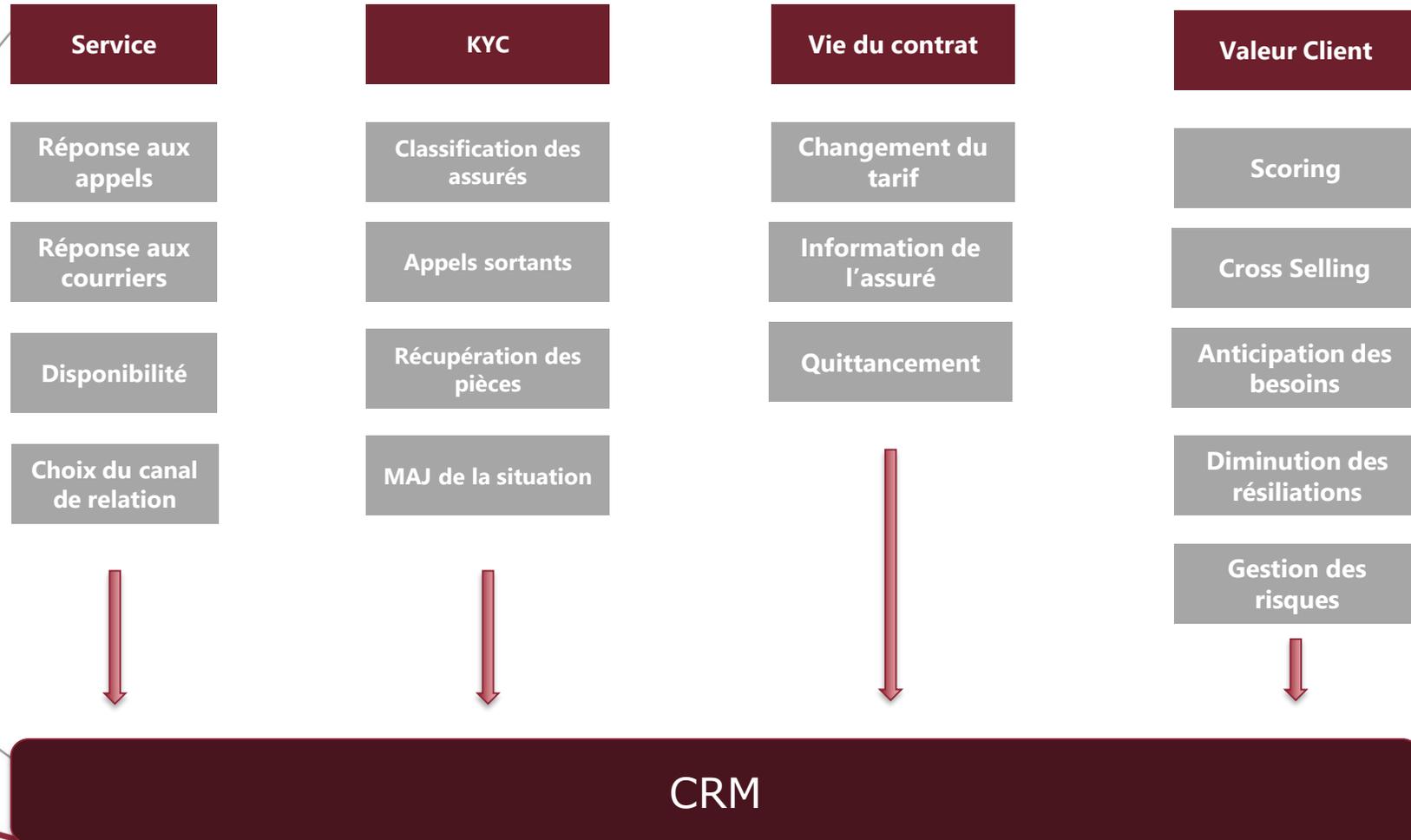
Au-delà du service client

Clarté

Proactivité



La gestion des contrats



Service

Analyse automatique
de mails et routage

ChatBot

KYC

Analyse des
mouvements

Récupération et
analyse des pièces

Vie du contrat

Individualisation
des majorations

Valeur Client

Vision client 360 °

Scores d'appétence

Scores de fragilité

Machine Learning

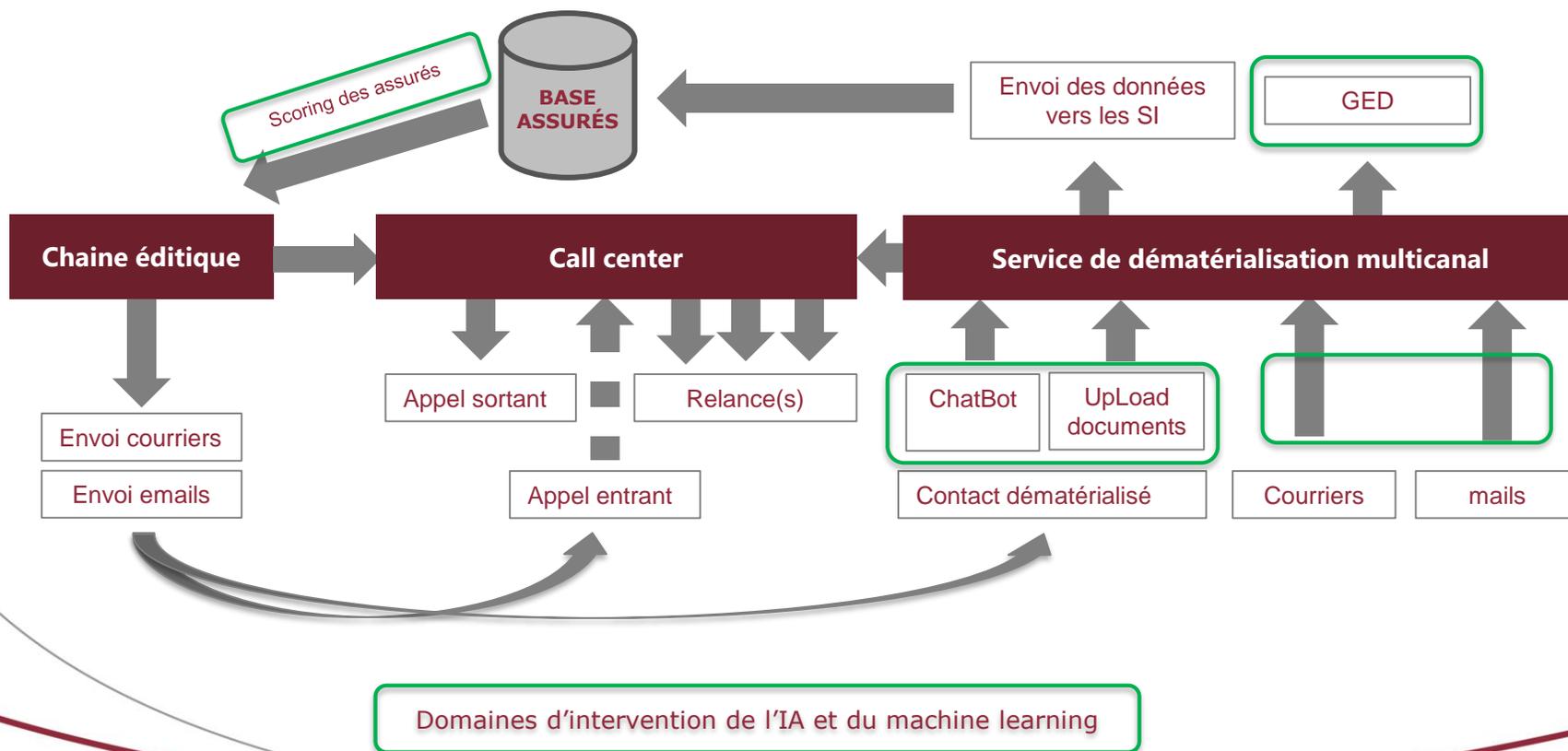
ChatBot

Langage naturel

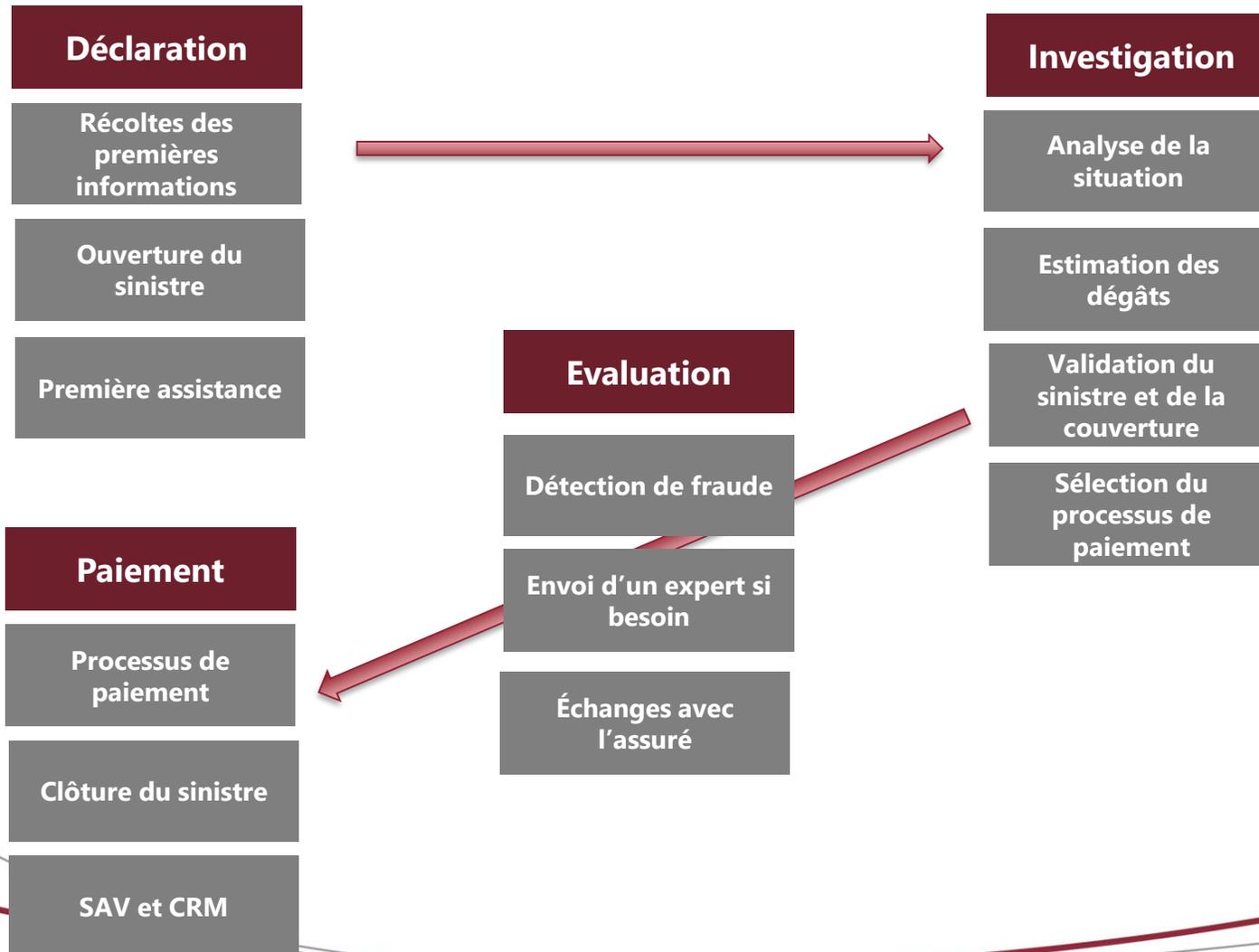
Reconnaissance
d'images

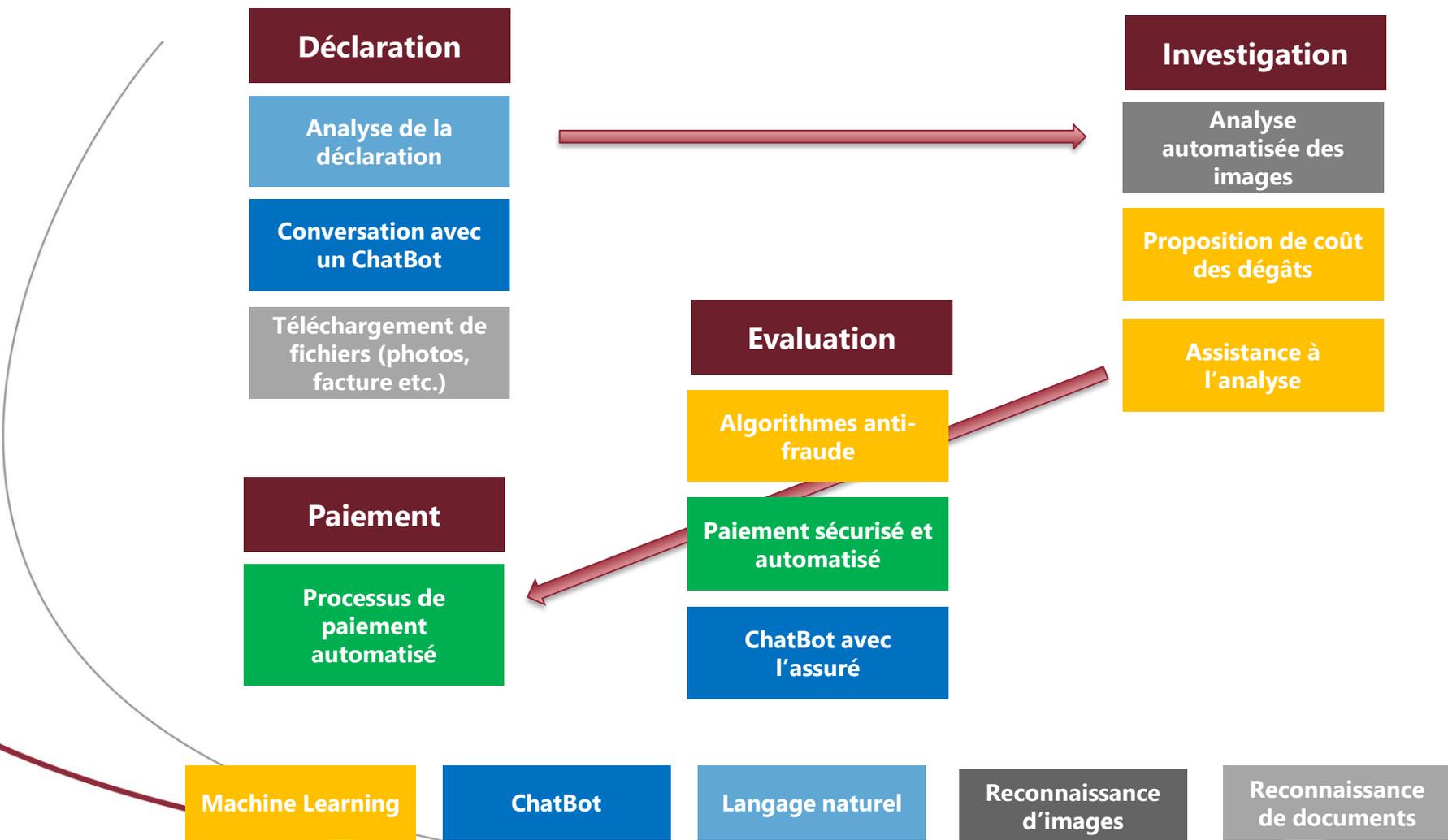
Reconnaissance
de documents

La mise à jour de la connaissance des assurés (KYC)



La gestion des sinistres





Objectifs :

Accélérer traitement dossiers
Fluidifier le parcours client
Réduire les coûts d'expertise

1. Classification d'image
2. Identification des marques et modèles
3. Lecture de la plaque d'immatriculation
4. Identification de parties de véhicules
5. Détection précise de dommages



Synthèse et conclusion

La nécessité de l'explicabilité des résultats



Transparence
des modèles



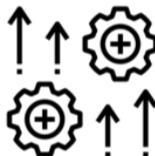
Générer de la
confiance



Respecter la
réglementation



Comprendre
la prise des
décisions



Améliorer la
performance
des modèles



Réduire les
biais éthiques
et moraux

De nouveaux risques à gérer

Risques juridiques

Utilisation des données

Risques opérationnels

Capacité de contrôle de la pertinence des propositions des outils

Explicabilité des résultats

Dégradation de la qualité des données

Risques d'image

Erreurs générant du mécontentement des assurés

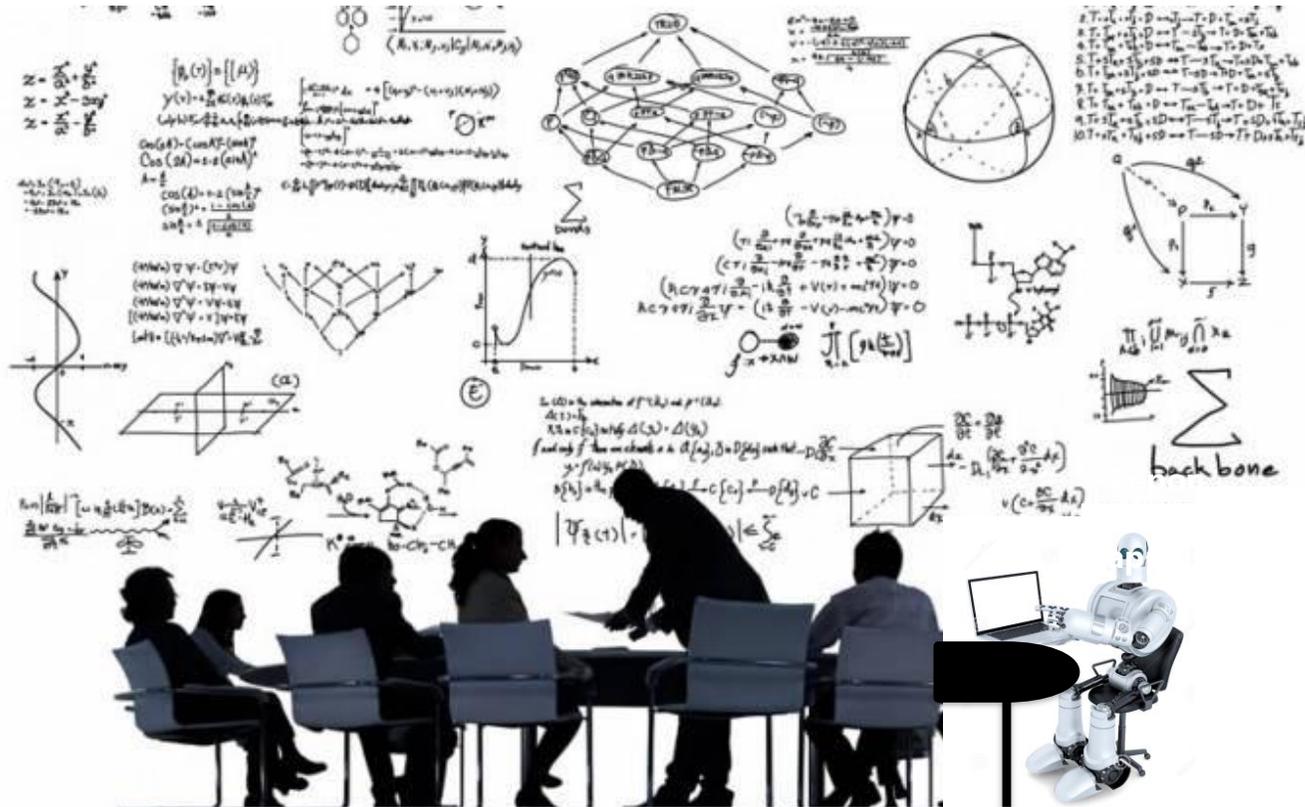
Communication et aversion du public sur l'IA

Risques IT

Intrusion dans les bases Big Data

Complexité des développements

Communication entre bases et outils très différents



L'actuaire augmenté

Questions / réponses

Nicolas Marescaux : 06 12 71 95 12

Marc Raymond : 07 66 51 64 69