



*Bienvenue!*

**ÉCOLE D'ÉTÉ FRANCOPHONE  
EN APPRENTISSAGE PROFOND**

---

**21-25 août 2017**



**IVADO**

HEC Montréal  
Polytechnique Montréal  
Université de Montréal



**MILA**

A complex network diagram with numerous nodes and connecting lines, some nodes are larger and black, others are smaller and grey.

# En pratique: **Mégadonnées**

Caroline Pernelle, Ing., PMP, PMI-ACP  
Conseillère aux partenariats industriels, IVADO



**IVADO**

HEC Montréal  
Polytechnique Montréal  
Université de Montréal



# IVADO

HEC Montréal  
Polytechnique Montréal  
Université de Montréal

CANADA  
EXCELLENCE  
RESEARCH  
CHAIR



**DATA SCIENCE  
FOR REAL-TIME  
DECISION-MAKING**



## CIRRELT

Département d'informatique  
et de recherche opérationnelle  
Faculté des arts et des sciences

Université   
de Montréal

# GERAD

## HEC MONTRÉAL

Department  
of Decision  
Sciences



## MILA



## POLYTECHNIQUE MONTRÉAL

DEPARTMENT OF MATHEMATICS  
AND INDUSTRIAL ENGINEERING



# Objectifs

**Rappels essentiels sur les mégadonnées**

**Stratégie & Définition des objectifs et des besoins**

**Outils & Processus**



# Rappels essentiels sur les mégadonnées

Regulating the internet giants

# The world's most valuable resource is no longer oil, but data

*The data economy demands a new approach to antitrust rules*



David Parkins

Advertisement

## Not all intelligence is artificial.



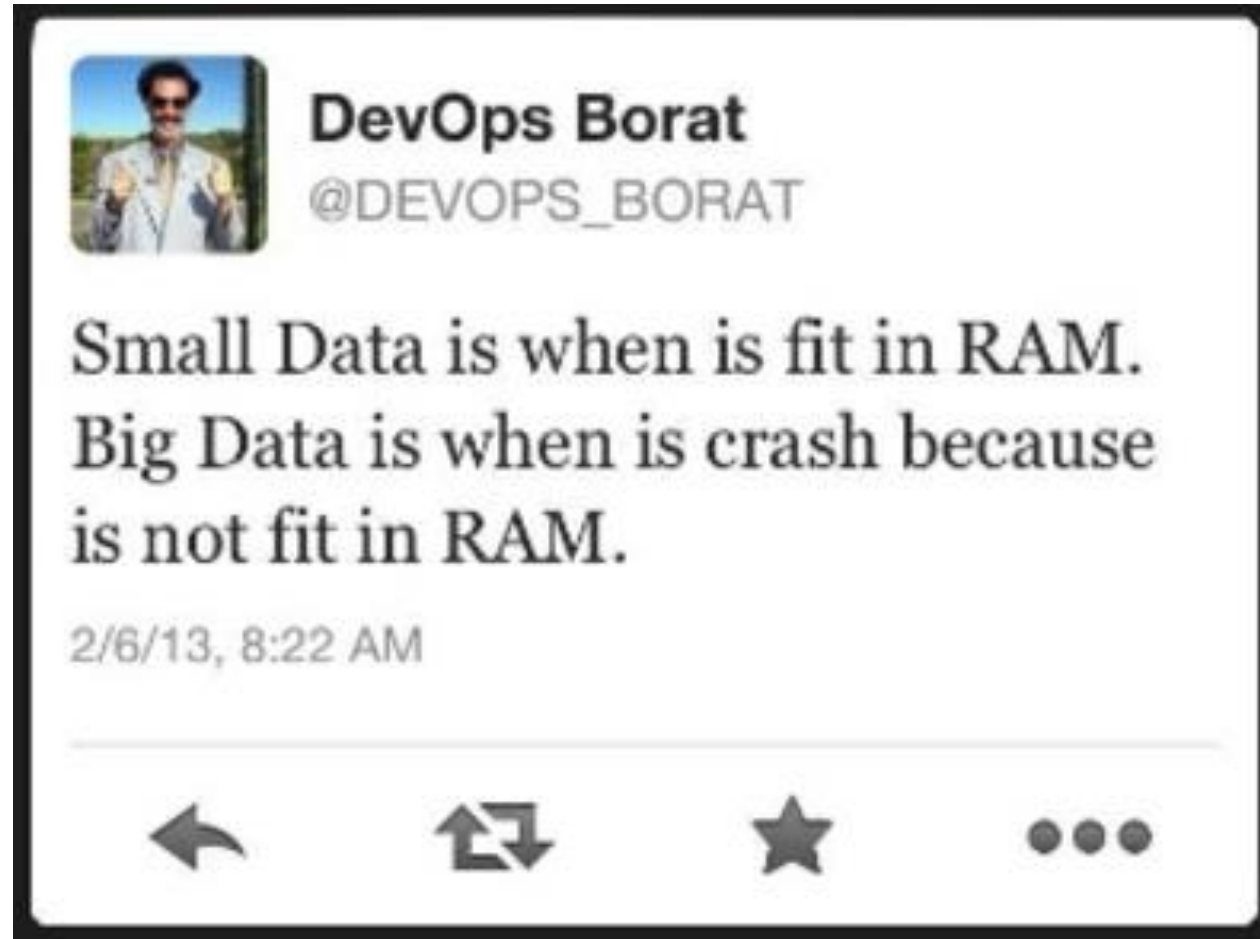
Enjoy 12 weeks' access for just \$12.

SUBSCRIBE ▶



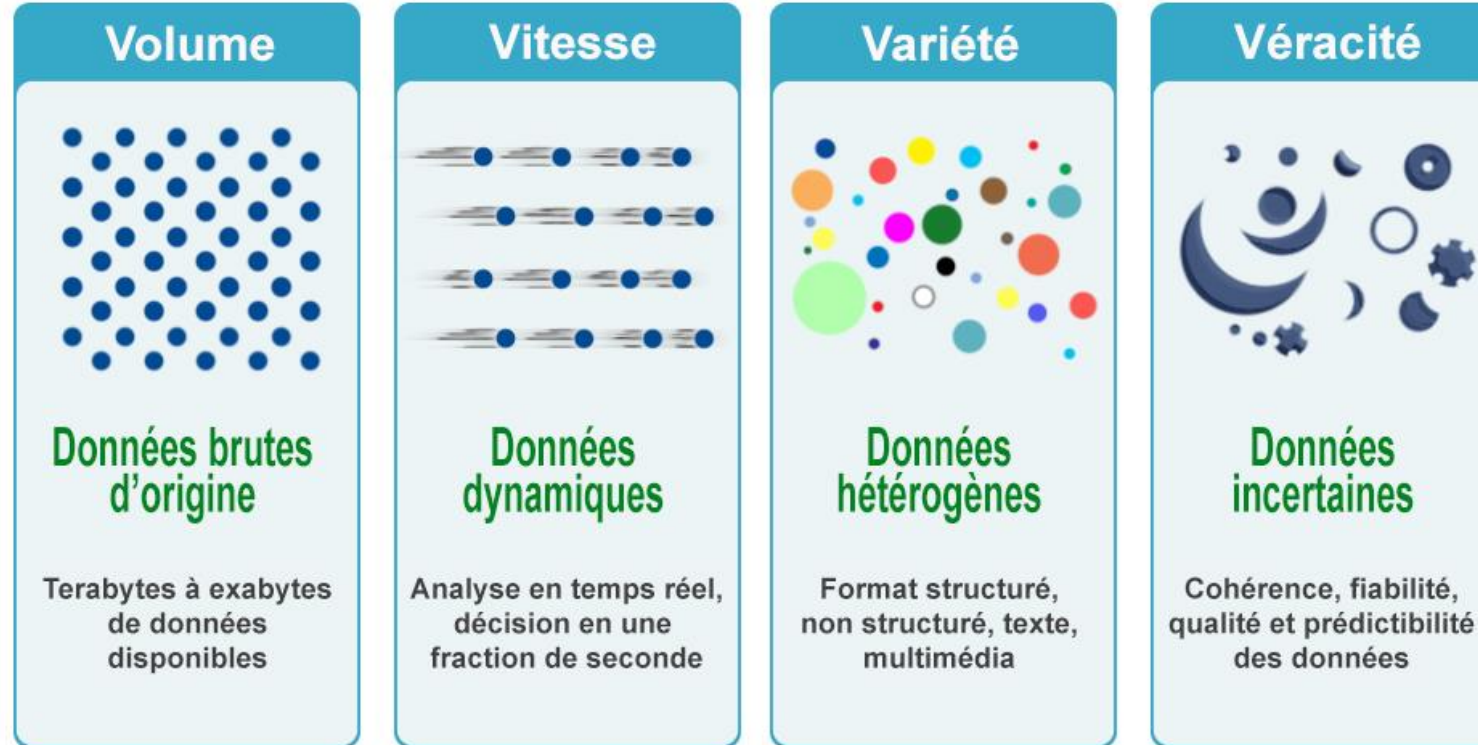
# IVADO

# Mais qu'est-ce que le Big Data?



**IVADO**

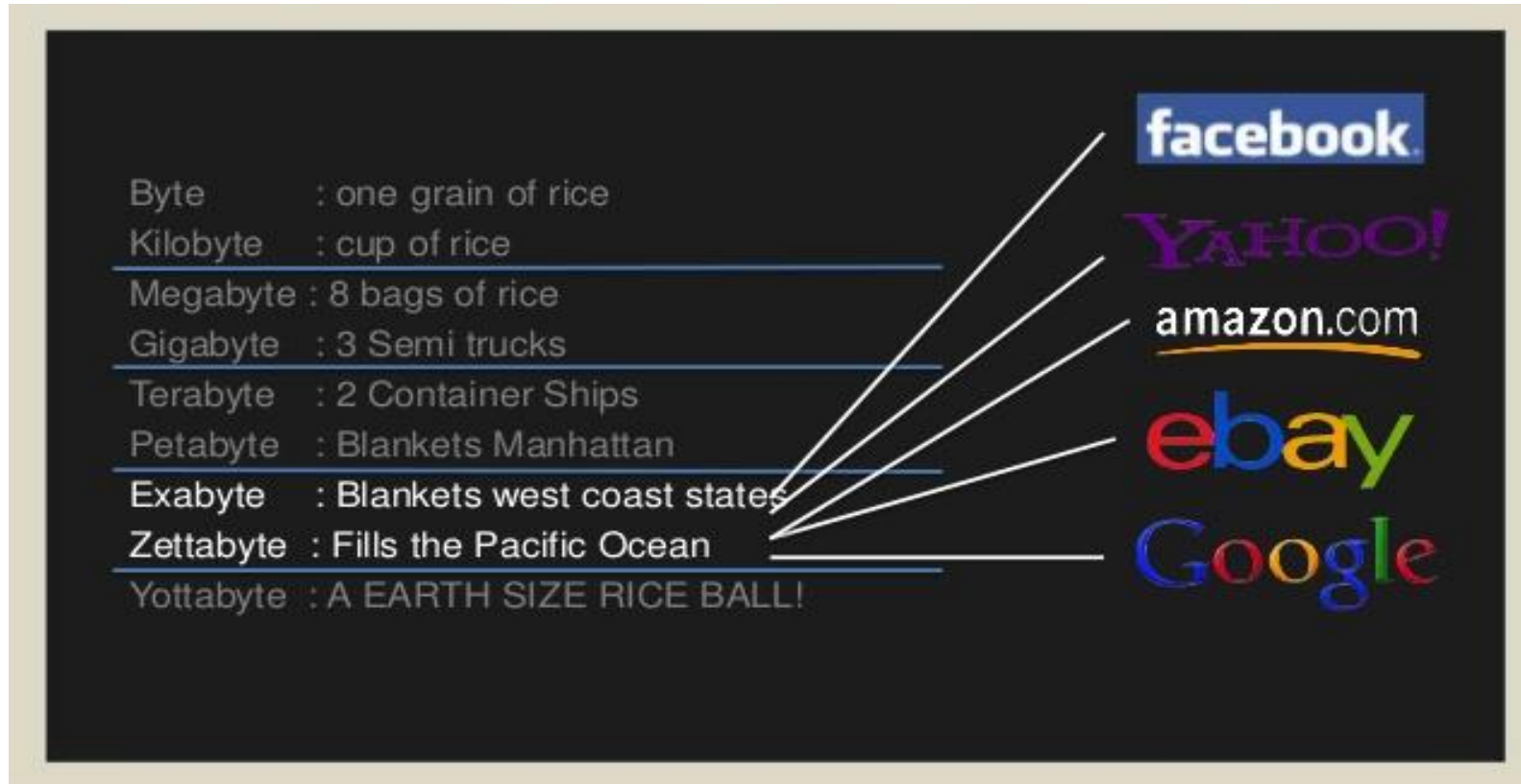
# Définition en 4 V



**IVADO**



# En grains de riz

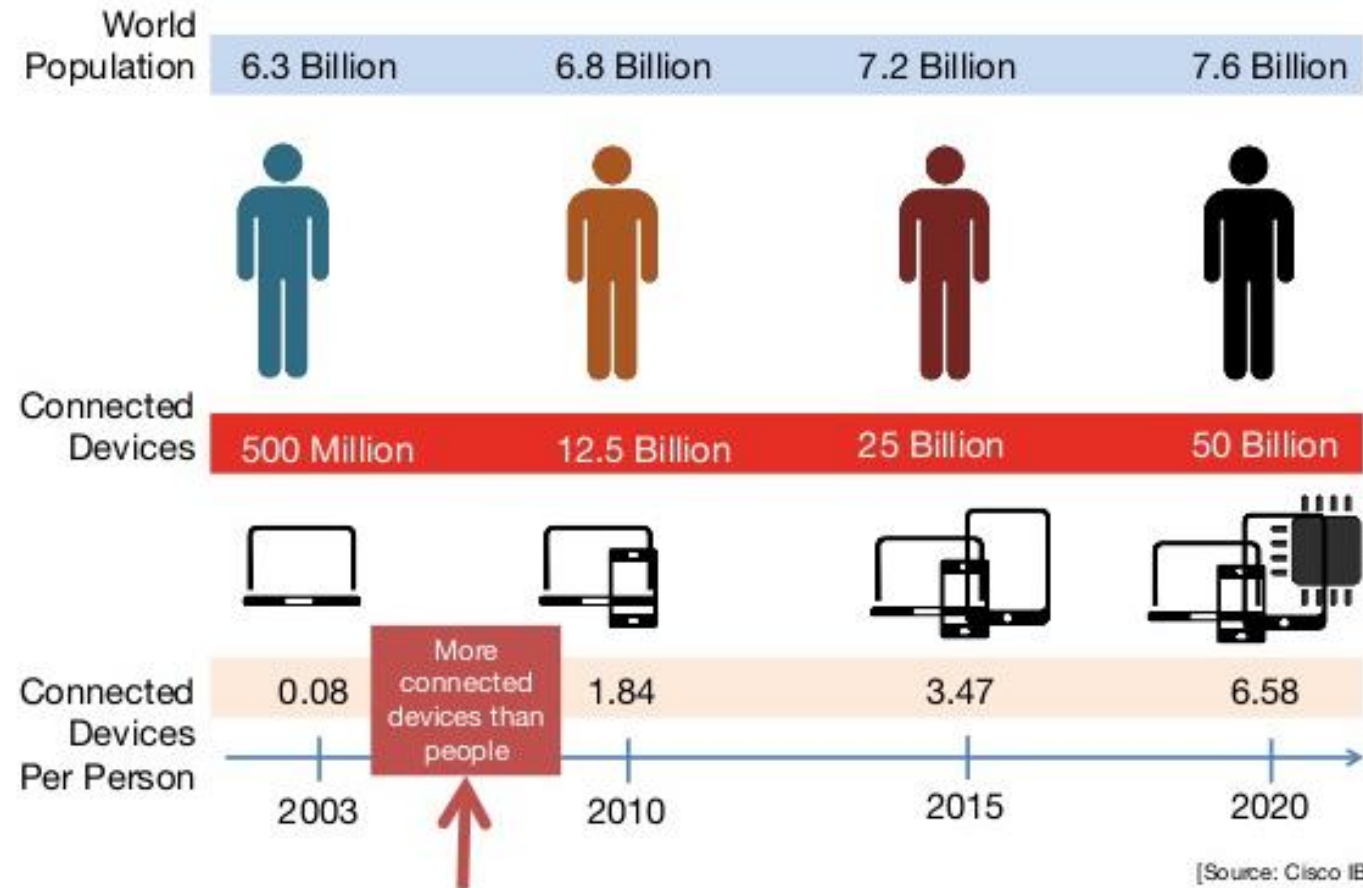


Comparison given by David Wellma  
<http://fr.slideshare.net/dwellman/what-is-big-data-24401517/>



IVADO

# Sources



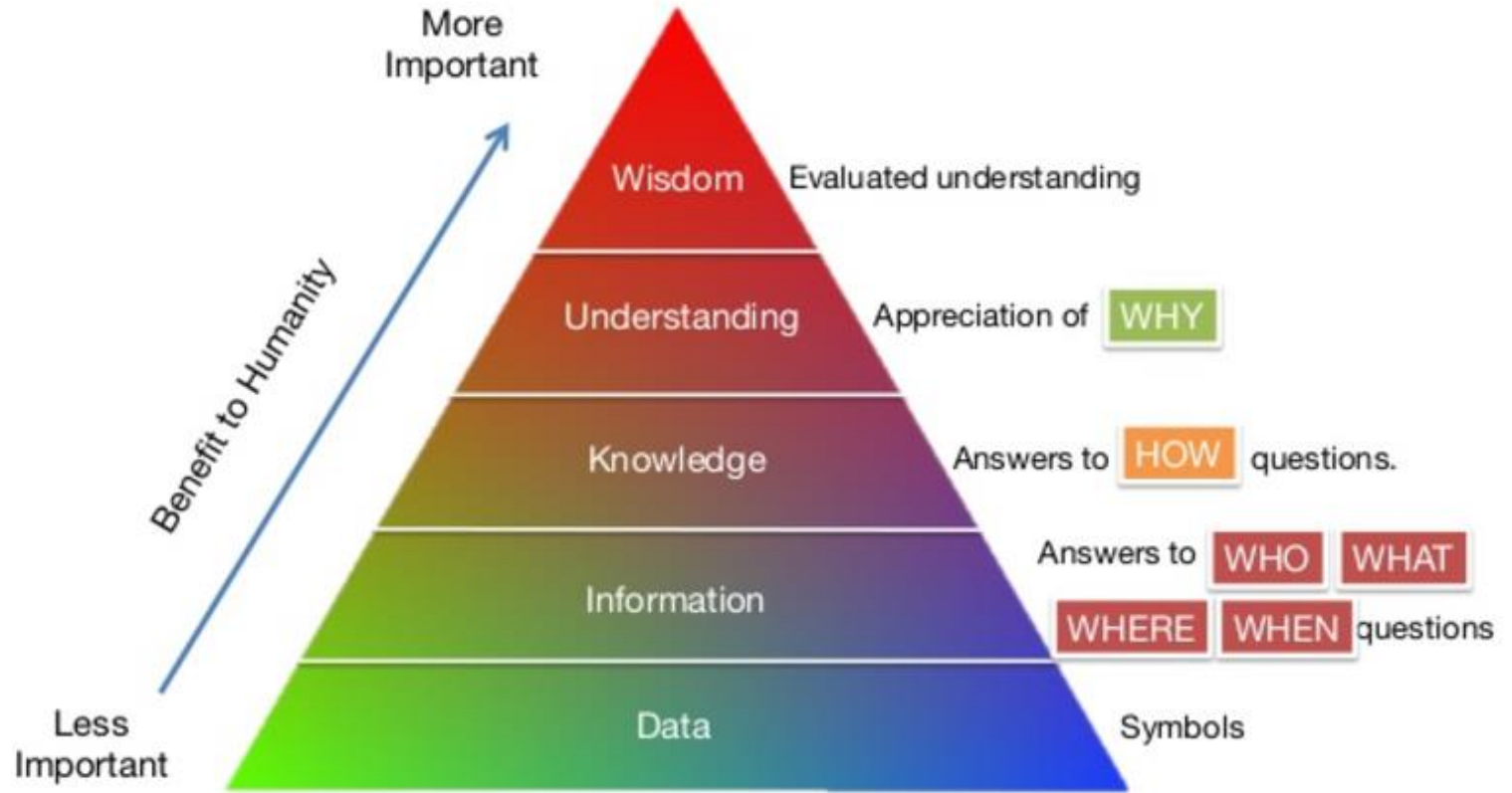
[Source: Cisco IBSG, April 2011]



IVADO

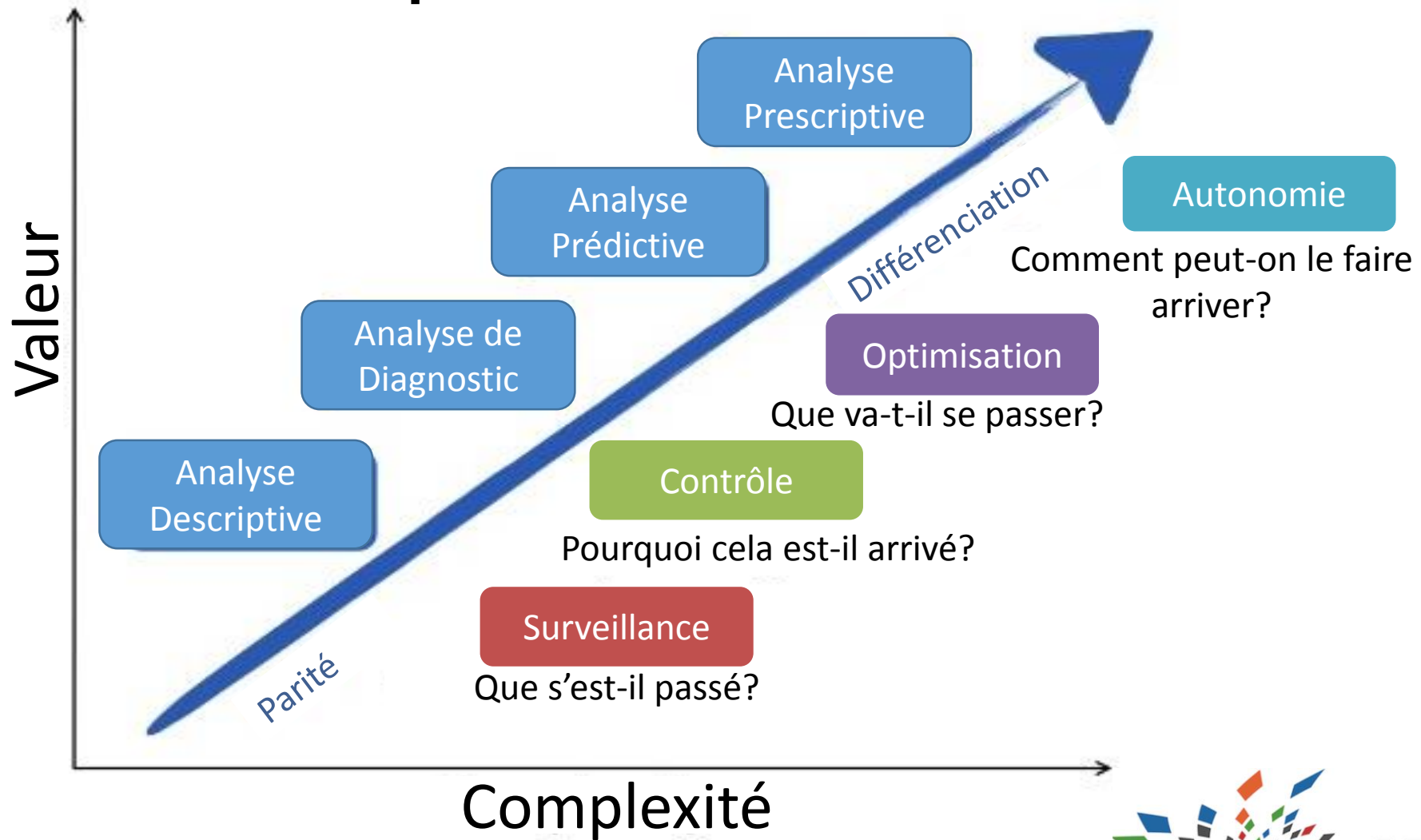
# Pourquoi avons-nous besoin de données?

- Ce n'est pas une question de taille, mais une question de valeur.




**IVADO**

# Valeur vs complexité



IVADO



*The ability to hear the signal  
from the noise is the key...*



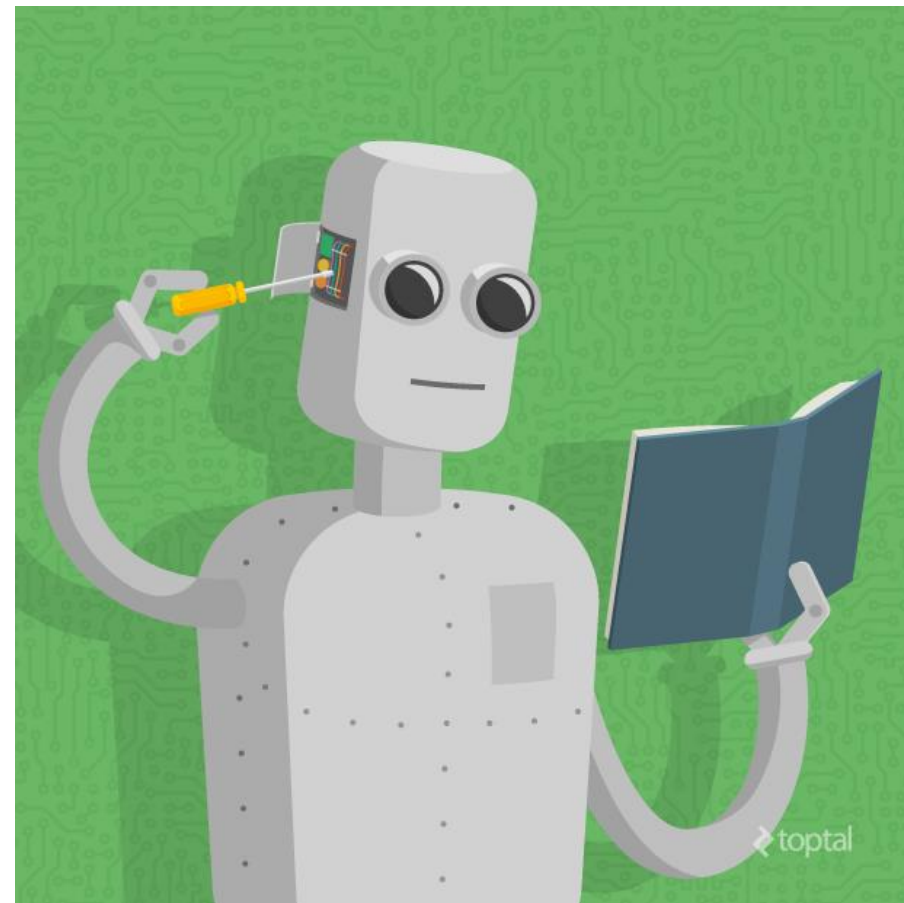
**IVADO**



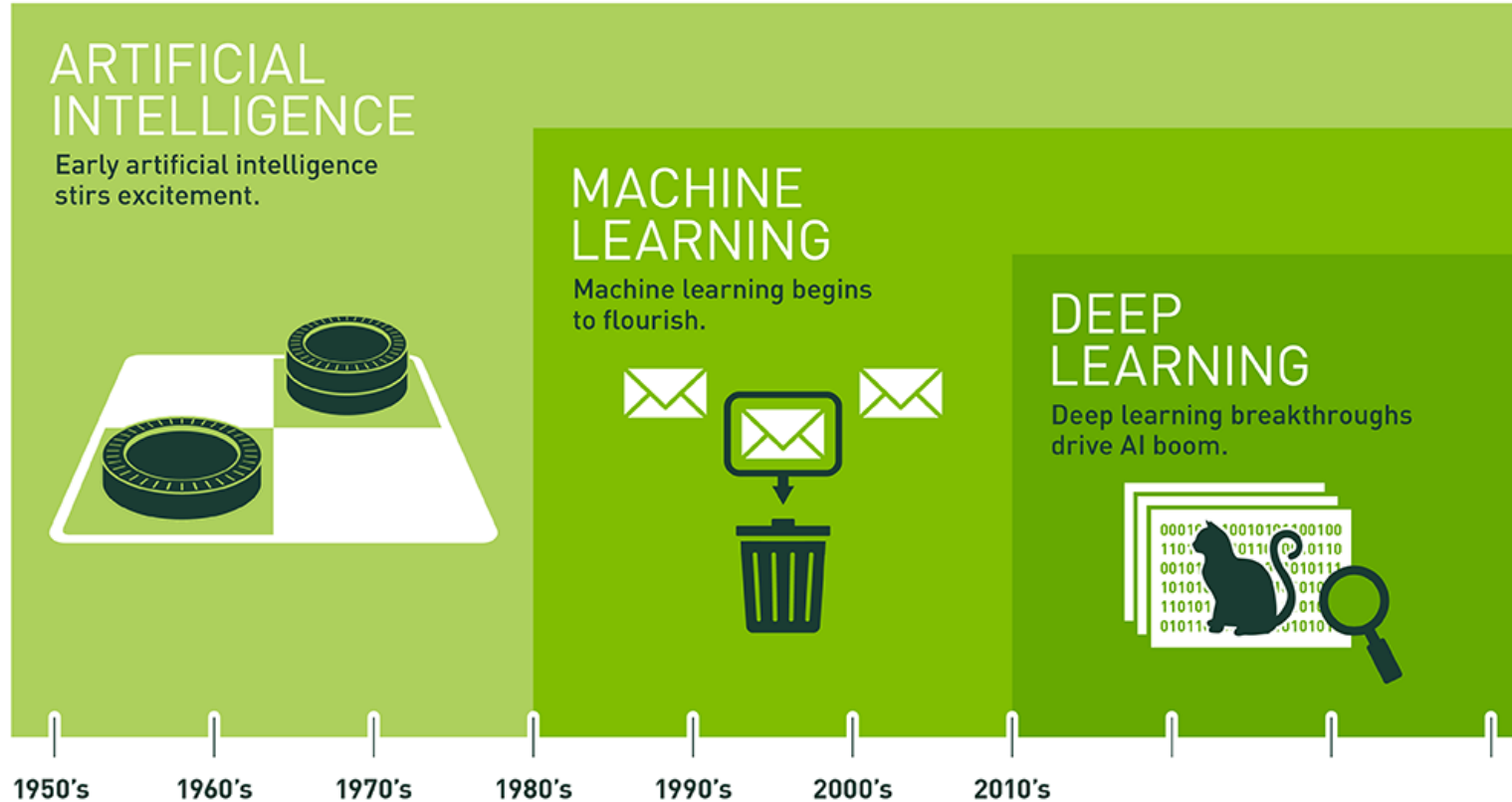
**Stratégie  
&  
Définition des objectifs et des besoins**

# Apprentissage Machine

- Capacité des ordinateurs d'apprendre à effectuer des tâches de classification ou de prédiction, sans intervention humaine



# Évolution temporelle



**IVADO**



# Mise en contexte

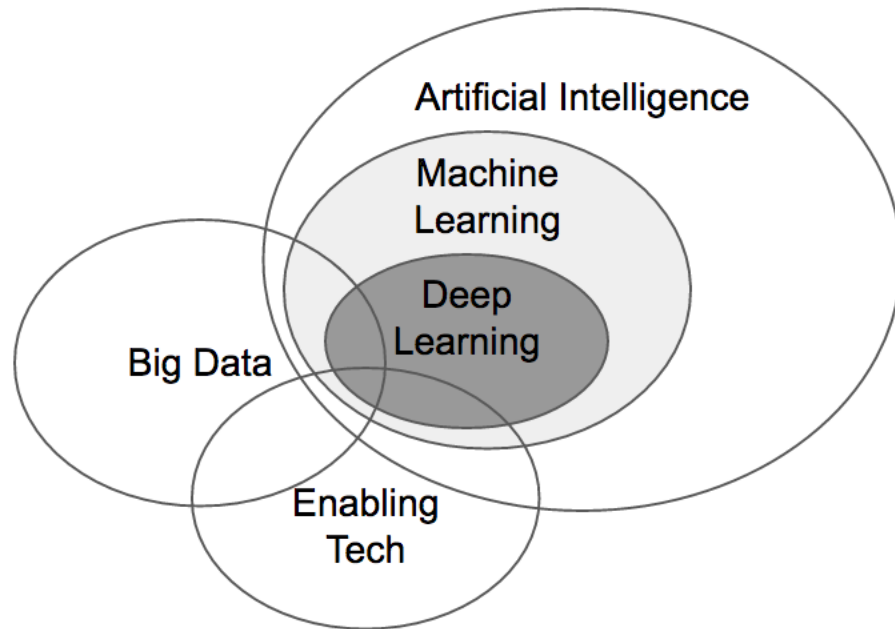
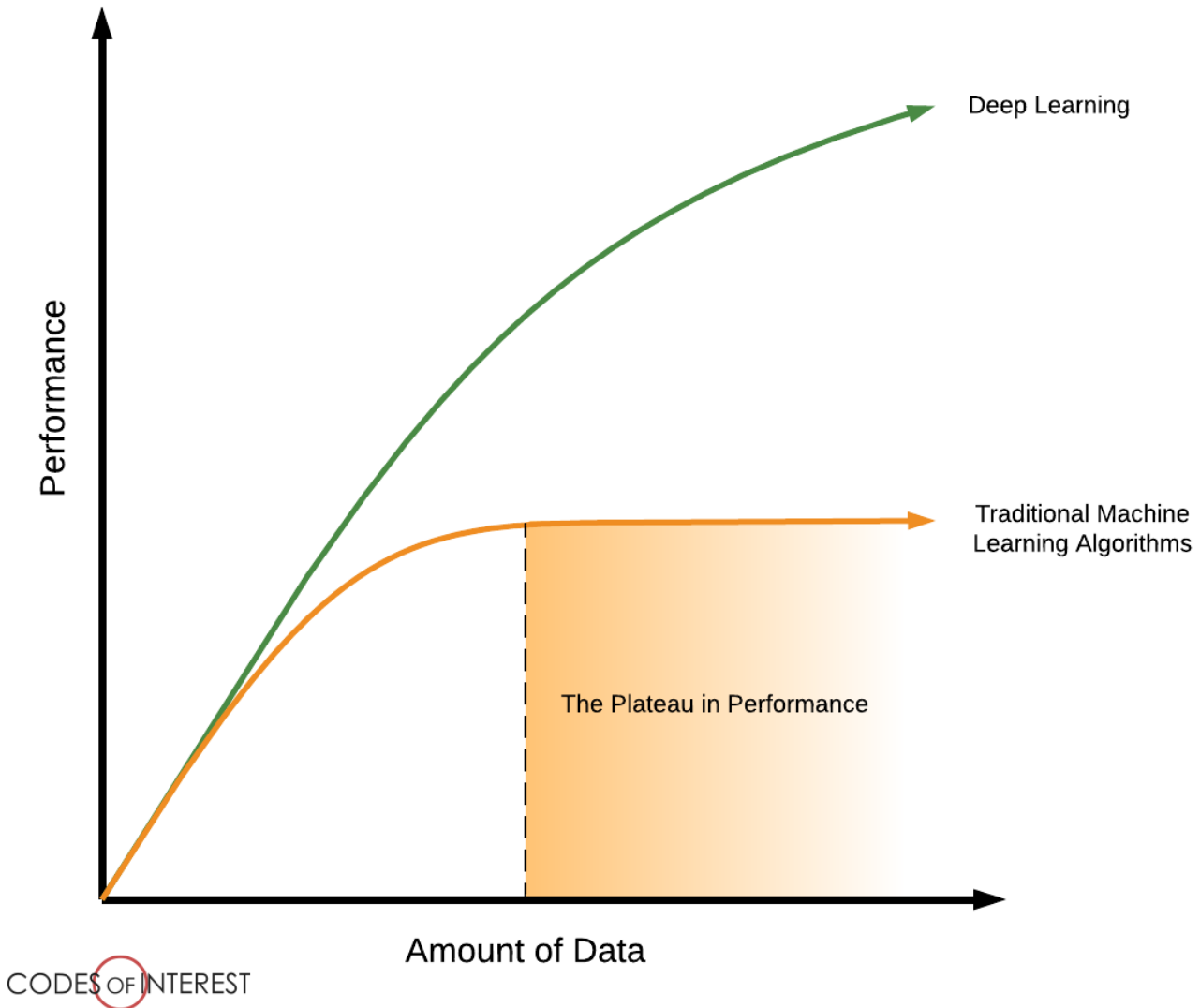


Diagramme de Venn

<http://www.gettingsmart.com/2017/03/the-technologies-reshaping-life-and-livelihood/>



# Avantages



<http://cs229.stanford.edu/materials/CS229-DeepLearning.pdf>



IVADO

# Recommandations d'approche

1. Définition d'objectifs clairs et gestion des attentes
2. Définition des métriques qui prouveront la valeur du projet
3. Stratégie pour la sélection des outils et du code déployé



# Sélection du bon projet

1. Projet à valeur ajoutée démontrable
2. Projet avec les bons champions
3. Projet à effet boule de neige
4. Projet à compétences transférables

**Breakthrough Big Data and Deep Learning in Today's Oil Industry**

<http://www.aapg.org/publications/blogs/learn/article/Articleid/37297/breakthrough-big-data-and-deep-learning-in-todays-oil-industry-interview-with-kamal-hami-eddine>

**AXA is using machine learning for insurance pricing optimization**

<https://cloud.google.com/blog/big-data/2017/03/using-machine-learning-for-insurance-pricing-optimization>

**Google's Inbox app offers three ready-made replies for many incoming emails**

<http://fortune.com/ai-artificial-intelligence-deep-machine-learning/>



# A quoi s'attendre?

1. Préparez-vous à du volume
2. Attendez-vous à de la variété
3. Choisissez un système véloce
4. Nettoyez et préparez les données
5. Pensez à la conformité et à la sécurité



**IVADO**

# Risques

- Mauvaise évaluation de la portée du projet
- Attentes mal gérées
- Incapacité de s'adapter
- Données mal explorées

Faits: 55% des projets avec des mégadonnées ne sont jamais finalisés

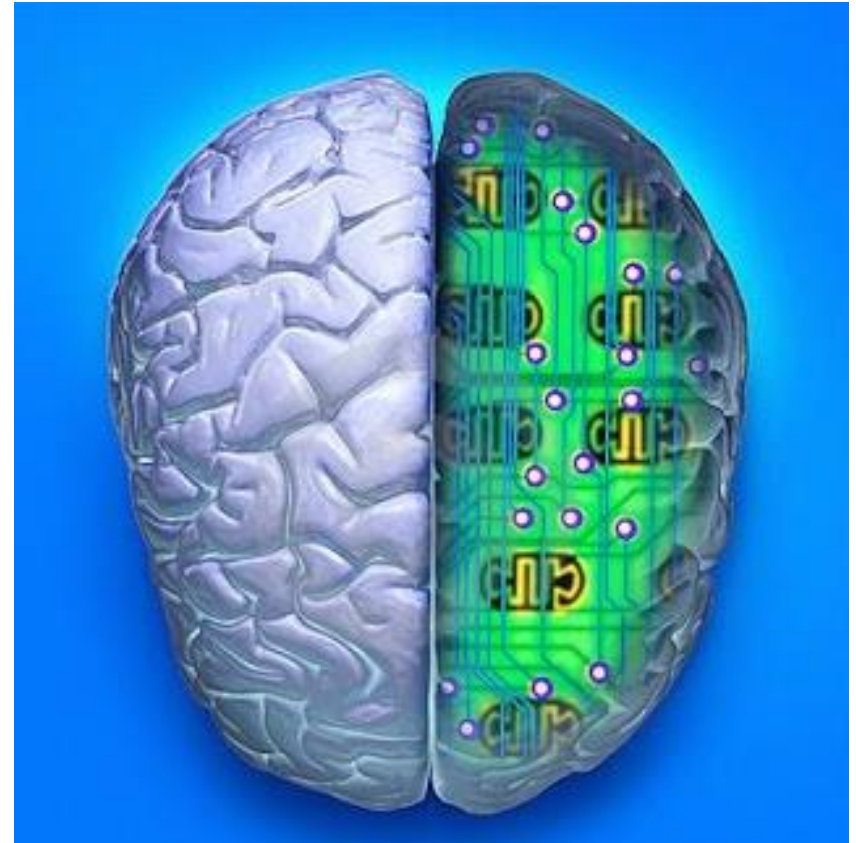
« Ça y est, j'ai des données. Je veux faire de l'IA. »



**IVADO**

# Limitations

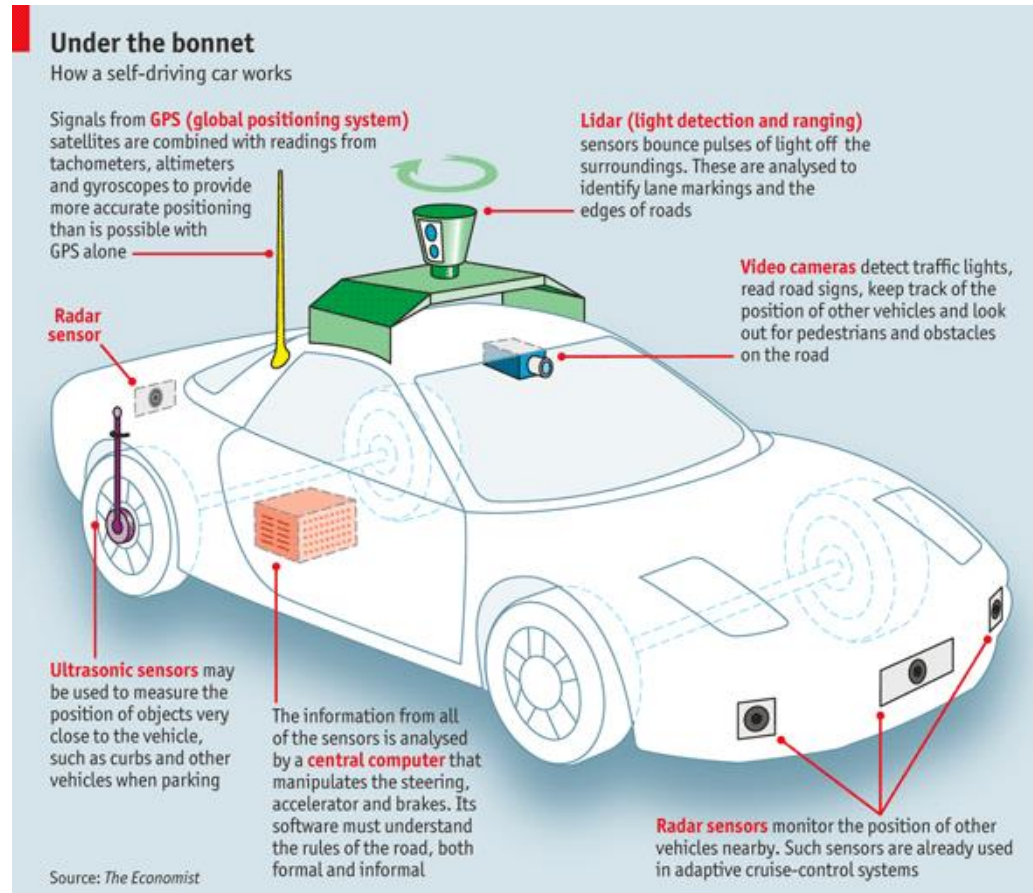
- Subjection
- Intuition
- Mise en contexte



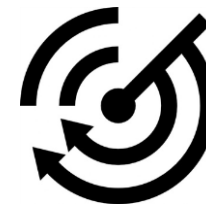
# Données – Cas de la voiture autonome



MOTION  
DETECTION



Source: *The Economist*



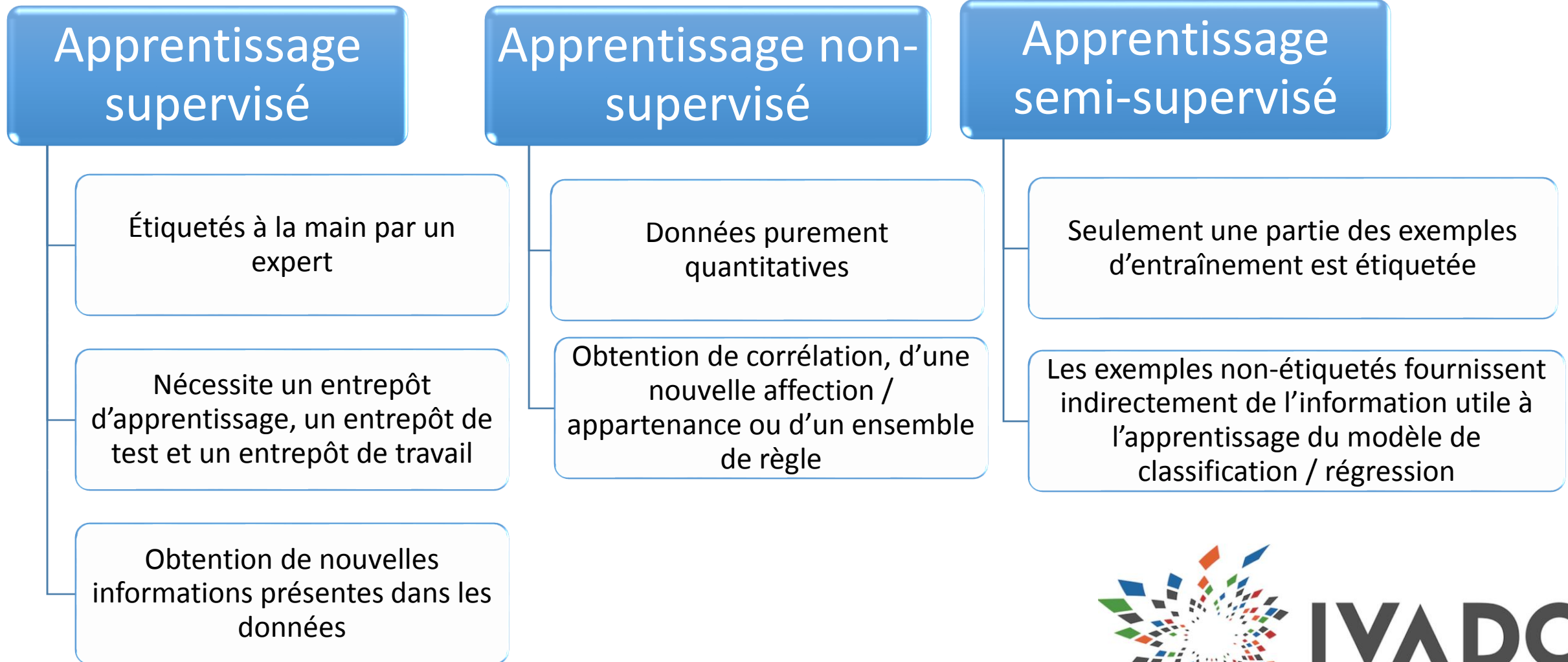
IVADO





## Outils & Processus

# Trois types d'apprentissage



# Les outils



Comment **générer** et **capturer** la valeur de la donnée?



Comment les nouvelles données (sensibles) sont-elles **utilisées** et **gérées**?



Comment **les canaux de communication** sont-ils redéfinis avec les partenaires?



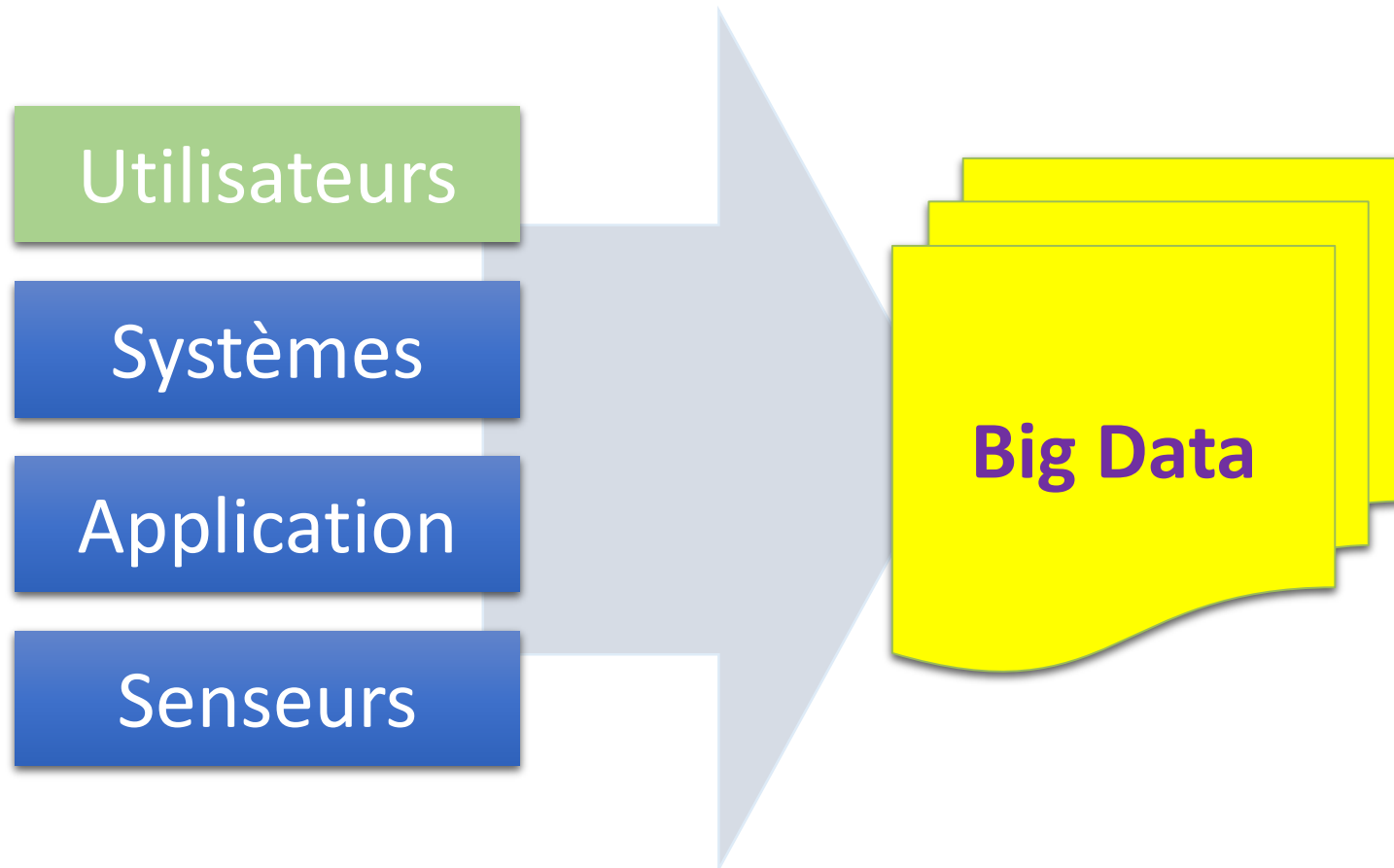
Quels **rôles** les compagnies devraient-elles jouer au fur et à mesure que les frontières des industries s'élargissent?

Fermé pour cause de ré-outillage



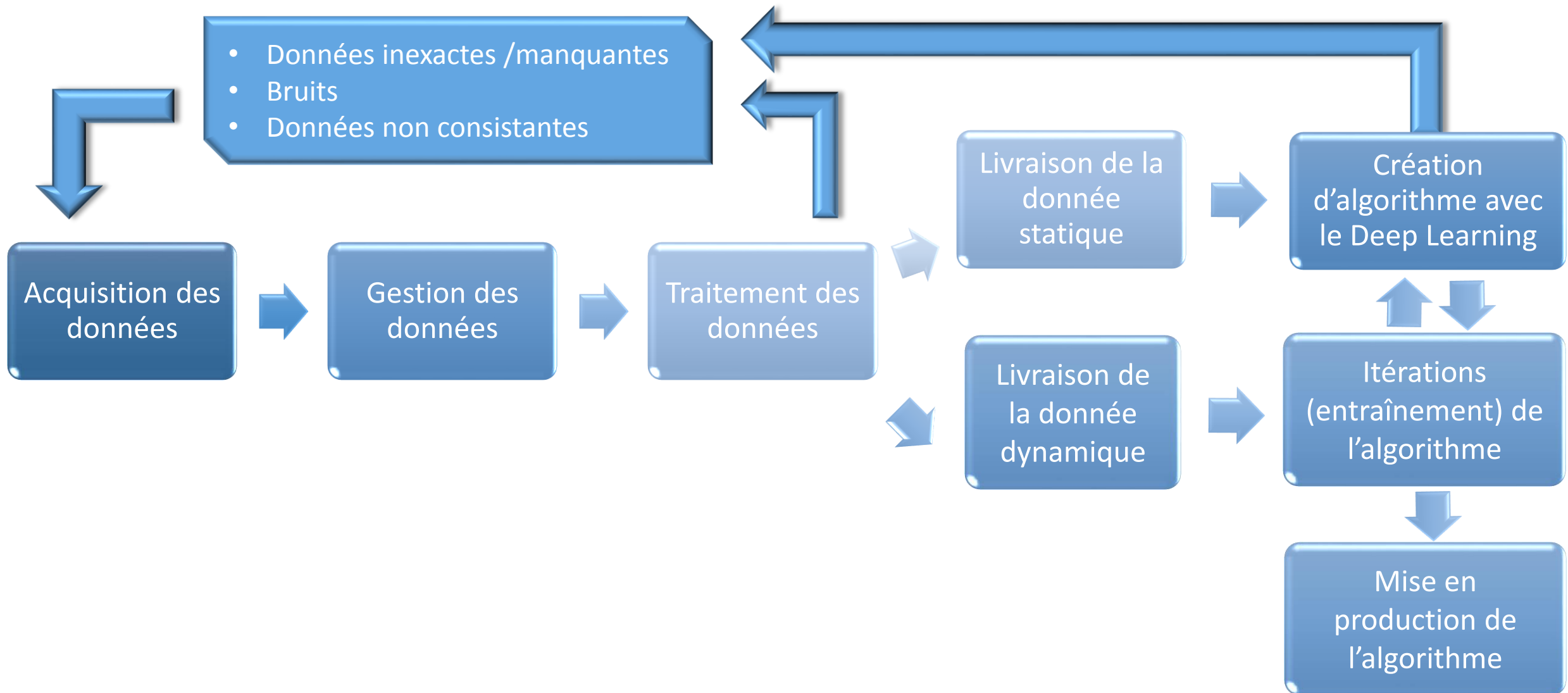
IVADO

# Les sources de données



**IVADO**

# Cycle de vie des données



# Exemple de données à ajuster



**IVADO**

# Conclusion

- Améliorer les performances des algorithmes grâce aux mégadonnées
- Commencer des projets d'apprentissage profond petits mais en ayant bien défini le besoin
- Mettre en place des systèmes d'exploitation de la donnée



**The more of your data I gather,  
the more I understand what it  
means to be *human***



**IVADO**





**IVADO**

HEC Montréal  
Polytechnique Montréal  
Université de Montréal

Questions?



# IVADO

HEC Montréal  
Polytechnique Montréal  
Université de Montréal

CANADA  
EXCELLENCE  
RESEARCH  
CHAIR



**DATA SCIENCE  
FOR REAL-TIME  
DECISION-MAKING**



## CIRRELT

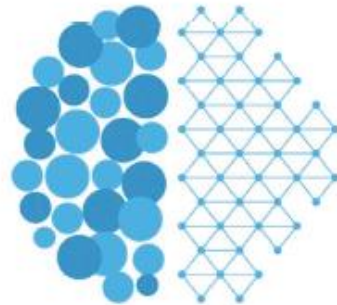
Département d'informatique  
et de recherche opérationnelle  
Faculté des arts et des sciences

Université   
de Montréal

# GERAD

## HEC MONTRÉAL

Department  
of Decision  
Sciences



## MILA



## POLYTECHNIQUE MONTRÉAL

DEPARTMENT OF MATHEMATICS  
AND INDUSTRIAL ENGINEERING