



## L'I.A. responsable @BNC

Éric Charton, Ph.D. & Dominique Payette, LL.M.

## **Cadre éthico-juridique**

- Législation anti-discrimination: respect du droit à l'égalité
  - Charte Canadienne
  - Charte Québécoise
  - Loi canadienne sur les droits de la personne
- GDPR vs Projet de loi 64
- **Devoirs éthiques de *fairness***
  - Corpus d'I.A. responsable
  - Codes de conduites et d'engagements sociaux

## **Application et mesures de gouvernance**

- Spécifique aux domaines de pratique
  - Domaine bancaire: services et produits généralement disponibles au grand public
- Volumétrie et risques systémiques
- *Fairness-by-design* – qualité des données et du développement
- Imputabilité, propriété et *monitoring*



## Exemple de risque dans un modèle:

- Modélisation par la résidence ou le segment socio-démographique
  - Captation de critères discriminatoire par relation implicite (population type d'une chaîne de restaurant, source de facturation – ex hôpital)
- 
- En pratique, les algorithmes fonctionnent par patterns et discriminent peu voir moins que les méthodes traditionnelles
    - *Ne pas entrer directement de paramètres interdits par la réglementation*



## Cadre éthico-juridique

- Législation *privacy*: décisions basées sur un traitement pleinement automatisés
  - PIPEDA: réforme 2.0 transparence algorithmique + consultation OPC
  - Projet de loi 64
  - GDPR
- Règlementation prudentielle et financière (OSFI et AMF)
- **Devoirs éthiques d'*explainability***
  - Corpus d'I.A. responsable
  - Codes de conduites et d'engagements sociaux

## Application et mesures de gouvernance

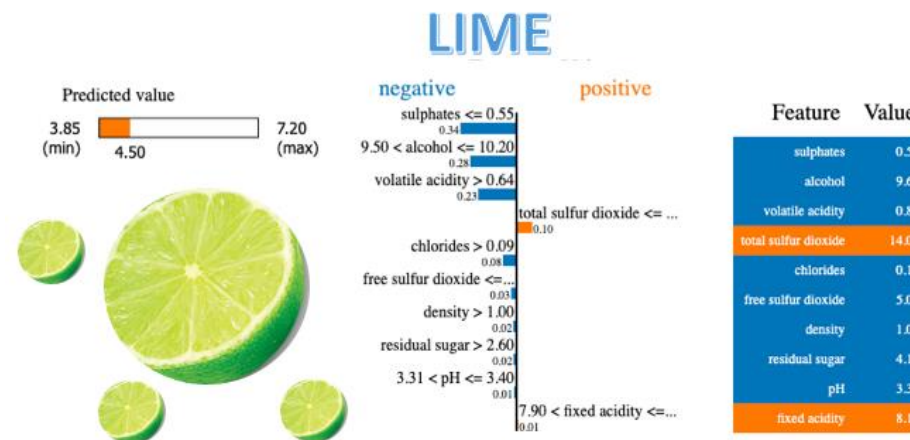
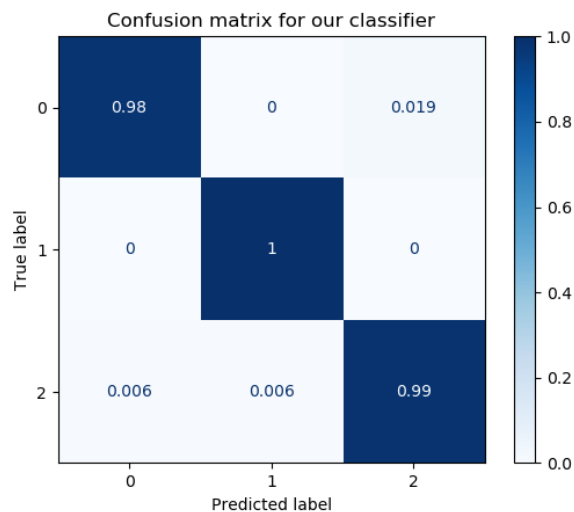
- *Privacy-by-design*:
  - Identification des renseignements personnels utilisés
  - Explication des facteurs, paramètres et logique ayant mené à la décision
- Domaine bancaire: validation réglementaire
- Canaux de révision et de contestation



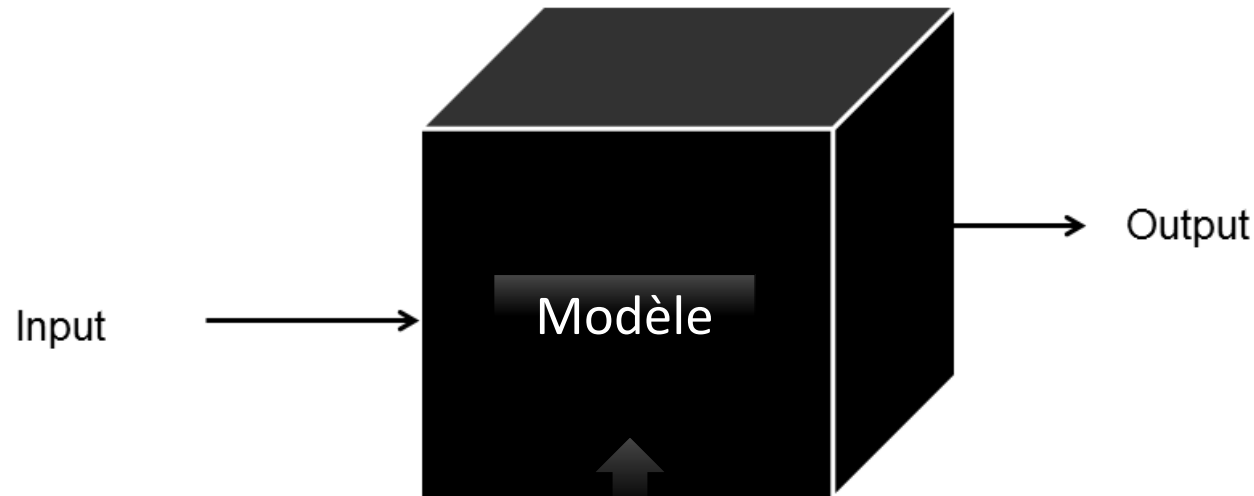
## L'explicabilité est une composante de la pratique AI

- Connaître les variables utilisées
  - Connaître les transformations appliquées
- Avoir une vue de leur impact sur les performances
  - Mesures conformes à l'état de l'art
- Avoir une vue précise du comportement des modèles
- Transparence du code
- Identification et documentation des données utilisées

L'ensemble des techniques de modélisation et conception de modèles est utilisée:



**En pratique, la blackbox en est rarement une (si tout est documenté)**



**Algorithme**

$$\begin{aligned} \text{maximize } f(c_1 \dots c_n) &= \sum_{i=1}^n c_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i c_i (\varphi(\vec{x}_i) \cdot \varphi(\vec{x}_j)) y_j c_j \\ &= \sum_{i=1}^n c_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i c_i k(\vec{x}_i, \vec{x}_j) y_j c_j \end{aligned}$$

**Les méthodes mathématiques et algorithmiques qui ont servi à construire le modèle sont connues, et les phénomènes identifiés ...**

## Cadre éthico-juridique

- Législation *privacy*: divulgation adéquate aux individus
  - PIPEDA: principe de transparence + consultation OPC
  - Projet de loi 64: divulgation des traitements pleinement automatisés
  - GDPR
- Principe général de Transparence dans PIPEDA
- **Devoirs éthiques de *transparency***
  - Corpus d'I.A. responsable
  - Codes de conduites et d'engagements sociaux

## Application et mesures de gouvernance

- Canaux de *opt-out* des décisions automatisées
- Mise en production : l'importance de l'aspect *business practice*
- HITL/HOTL: transfert d'information et influence



## - La transparence est rendue simple si :

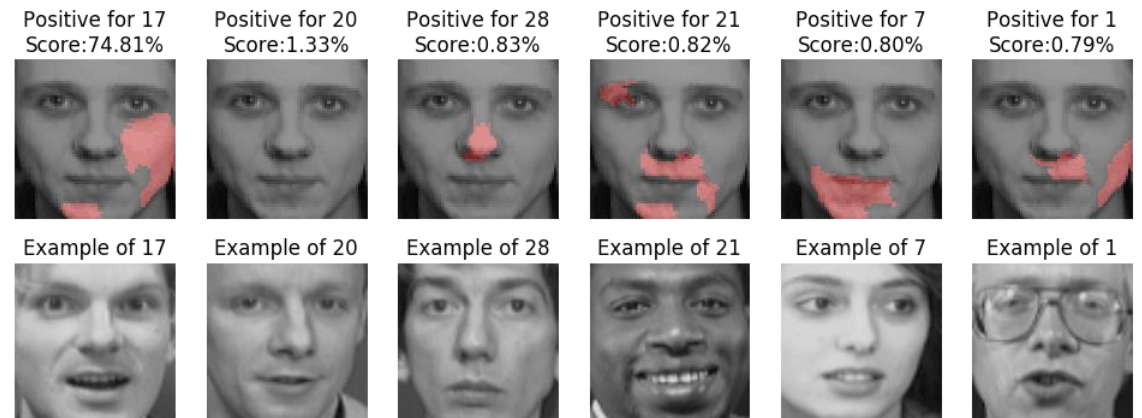
- Les modèles sont correctement documentés
- Les sources de données sont claires et correctement documentées
- Le taux d'erreur du modèle est expliqué et fait l'objet de prise en charge
- La complexité du processus d'affaire de bout en bout est prise en compte
  - (le modèle n'est pas tout)

## Certaines applications de l'IA sont plus difficiles à rendre transparentes

- Reconnaissance de locuteur, transcription vocale
- Reconnaissance faciale
- Modèles à forts besoins de calculs (1)

### • Ces modèles sont difficiles à interpréter:

- en raison des transformations des paramètres (ex: convolutions)
- En raison du manque de reproductibilité des techniques (Auto ML)
- En raison de la taille des espaces de représentation des paramètres (plusieurs milliers de dimensions)



(1) <https://arxiv.org/abs/1907.12363> A comparison of Deep Learning performances with other machine learning algorithms on credit scoring unbalanced data

