

WEBINAIRE

# REPRODUCTIBILITÉ EN APPRENTISSAGE AUTOMATIQUE

30 OCTOBRE 2020



## OBJECTIFS DE LA PRÉSENTATION

- Sensibiliser sur les enjeux de la reproductibilité.
  - Inciter l'intégration des solutions permettant une meilleure reproductibilité dans vos solutions d'affaires ou académiques.
  - Améliorer votre productivité.
- 



## VOTRE CONFÉRENCIER



**DAVID BEAUCHEMIN**

Candidat au doctorat

Département d'informa-  
tique et de génie logiciel

- Introduit à la recherche reproductible en 2016 (R Markdown et Git)
- Participation à REPROLANG de la conférence LREC [?]
- Membre actif dans le développement d'une librairie facilitant la reproductibilité ([Poutyne](#))

# Introduction



## C'EST QUOI LA REPRODUCTIBILITÉ?

La reproductibilité est le principe qu'on ne peut tirer de conclusions que d'un événement bien décrit, qui est apparu plusieurs fois, provoqué par des **personnes différentes**.

Toutefois, ont utilise souvent ce terme pour spécifiquement désigner la **réplicabilité**. Soit la réplication (reproduction) des résultats d'un article dans des environnements pas (toujours) différents [?, ?].





## AU MENU



Répliquer





## AU MENU



Répliquer



Jeux de données





## AU MENU



Répliquer



Jeux de données



Procédure  
d'entraînement



## AU MENU



Répliquer



Jeux de données



Procédure  
d'entraînement



Code



## POURQUOI S'Y INTÉRESSER ?

70 %<sup>1</sup>

---

1. [?]





## POURQUOI S'Y INTÉRESSER ?

50 %<sup>1</sup>

---

1. [?]





## POURQUOI S'Y INTÉRESSER ?

40 %<sup>2</sup>



## MOTIVATION



Réutilisation





## MOTIVATION



Réutilisation

Productivité





## MOTIVATION



Réutilisation



Productivité



Transfert





## MOTIVATION



Réutilisation



Productivité



Transfert



Se faire connaître



## **Les barrières à la réplicabilité**



Non-disponibilité des  
données



Non-disponibilité des  
données



Mauvaise  
spécification



Non-disponibilité des  
données



Mauvaise  
spécification



Non-disponibilité du  
code



Non-disponibilité des données



Mauvaise spécification



Non-disponibilité du code Non-disponibilité des configurations



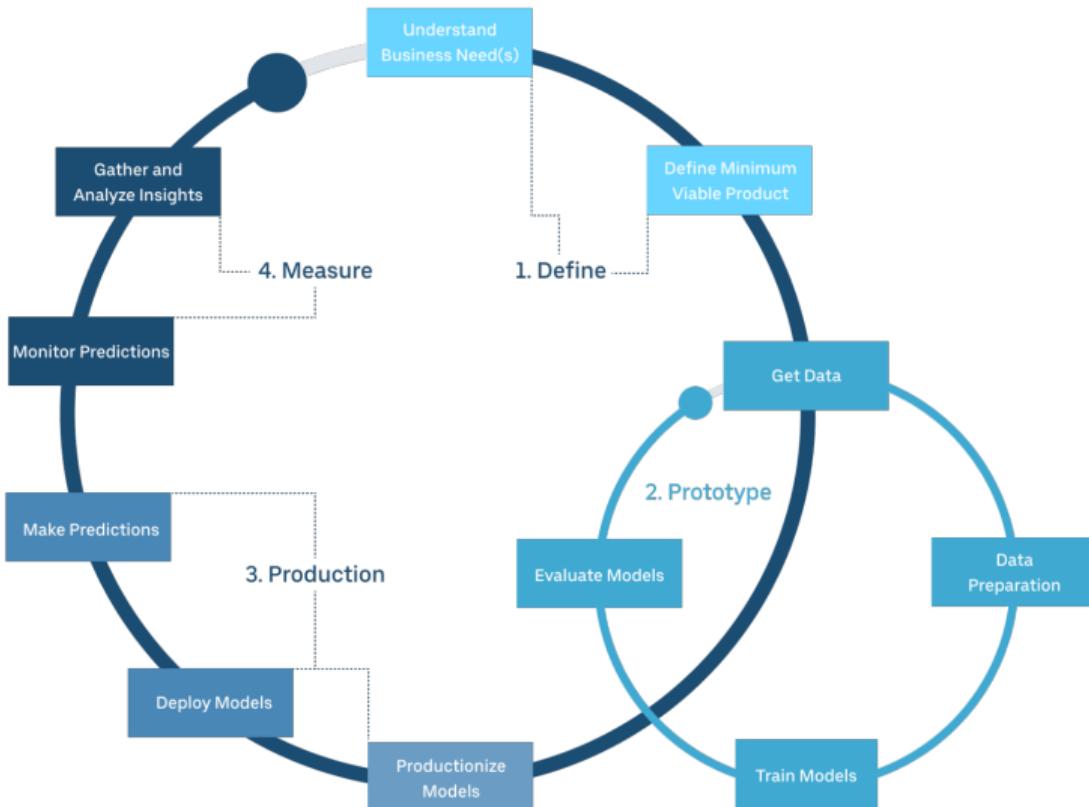
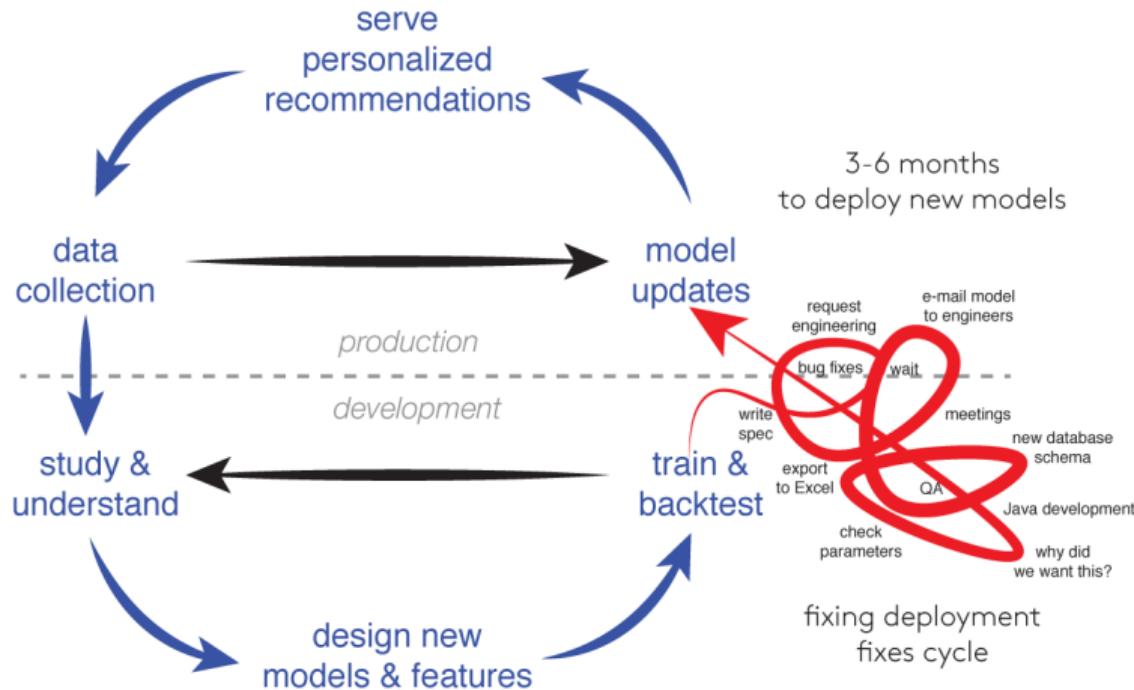
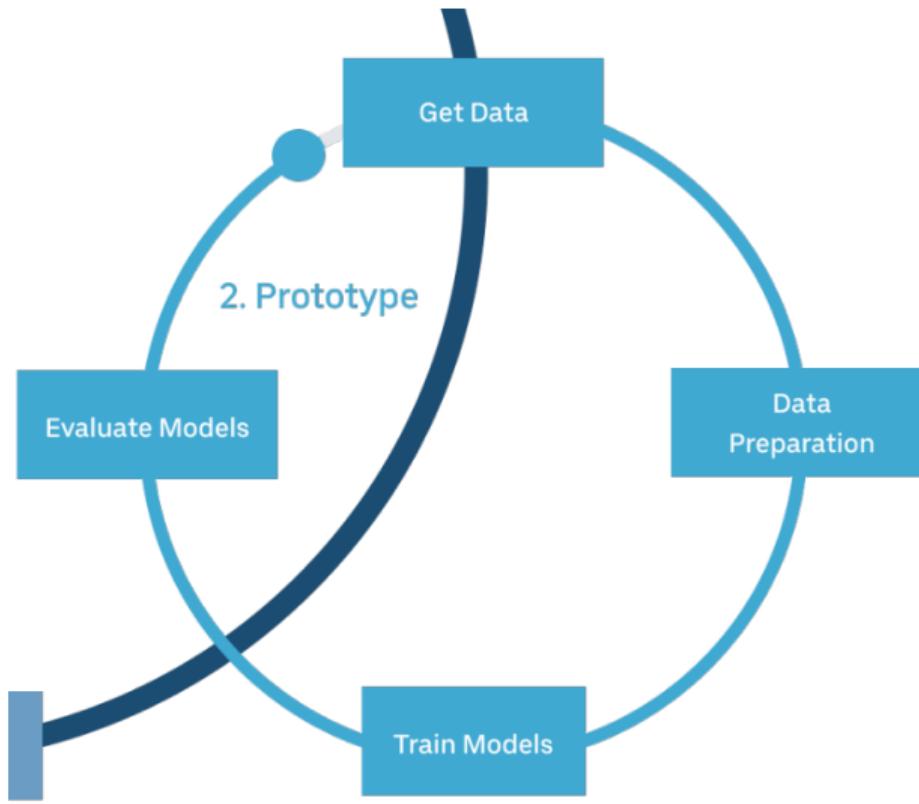


Figure 1 – From Uber Engineering



**Figure 2 – The need for Agile machine learning**

**OK, mais comment?**



**Figure 3 – From Uber Engineering**



## VERSION DES DONNÉES & ÉTAPES DE PRÉTRAITEMENT



Version





## VERSION DES DONNÉES & ÉTAPES DE PRÉTRAITEMENT



Version



Gestion des versions





## VERSION DES DONNÉES & ÉTAPES DE PRÉTRAITEMENT



Version



Gestion des versions



Étapes prétraitement





## VERSION DES DONNÉES & ÉTAPES DE PRÉTRAITEMENT



Data Version Control



Dask





CODE



Version





CODE



Version



Différence





CODE



Version



Différence



Divergences





CODE

git

Git



GitHub



GitLab



Bitbucket



## DÉVELOPPEMENT DES MODÈLES



Réinventer





## DÉVELOPPEMENT DES MODÈLES



Réinventer



Simplification





## DÉVELOPPEMENT DES MODÈLES



Réinventer



Simplification



Facilite





## DÉVELOPPEMENT DES MODÈLES



Poutyne



PyTorch  
Lightning



Scikit-learn



Gensim



Allen NLP



## ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS



Version de  
l'entraînement





## ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS



Version de  
l'entraînement



Résultats





## ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS



Version de  
l'entraînement



Résultats



Visualisation





## ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS



Version de  
l'entraînement



Résultats



Visualisation



Erreurs  
d'entraînement





## ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS



MLflow



Hydra



Sacred



Notif



# RAPPORT ET ANALYSE DES RÉSULTATS



Tableau des résultats





# RAPPORT ET ANALYSE DES RÉSULTATS



Tableau des résultats

Mise à jour





## RAPPORT ET ANALYSE DES RÉSULTATS



Tableau des résultats



Mise à jour



Visualisation configuration



# RAPPORT ET ANALYSE DES RÉSULTATS



Python2LaTeX



TensorBoard



Markdown



## ENVIRONNEMENT



Différents environnements





## ENVIRONNEMENT



Différents environnements



Réutilisation





## ENVIRONNEMENT



Docker



kubernetes

Kubernetes



**La suite**



Itérations d'expérimentations



## POUR ALLER PLUS LOIN

- Clean code
  - Continuous Machine Learning
  - Faire des tests!
  - Writing Code for NLP Research [?]
  - SOLID
  - Cet article [?]
- 



## PÉRIODE DE QUESTIONS



# WEBINAIRE

# MERCI DE VOTRE ÉCOUTE !



## REFERENCES i

