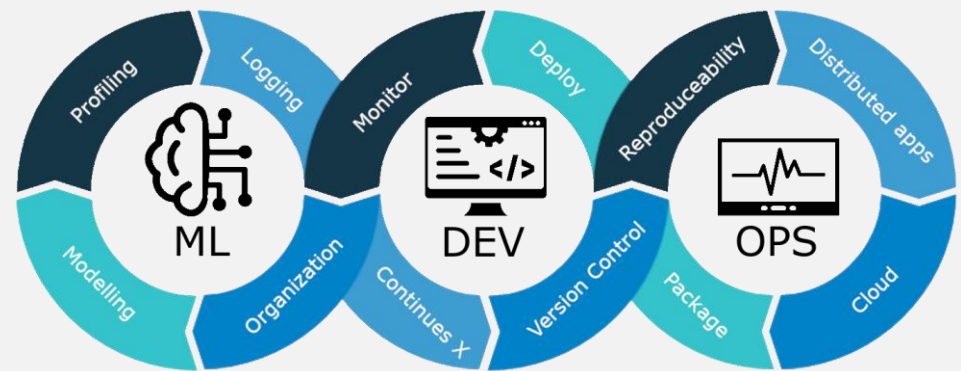


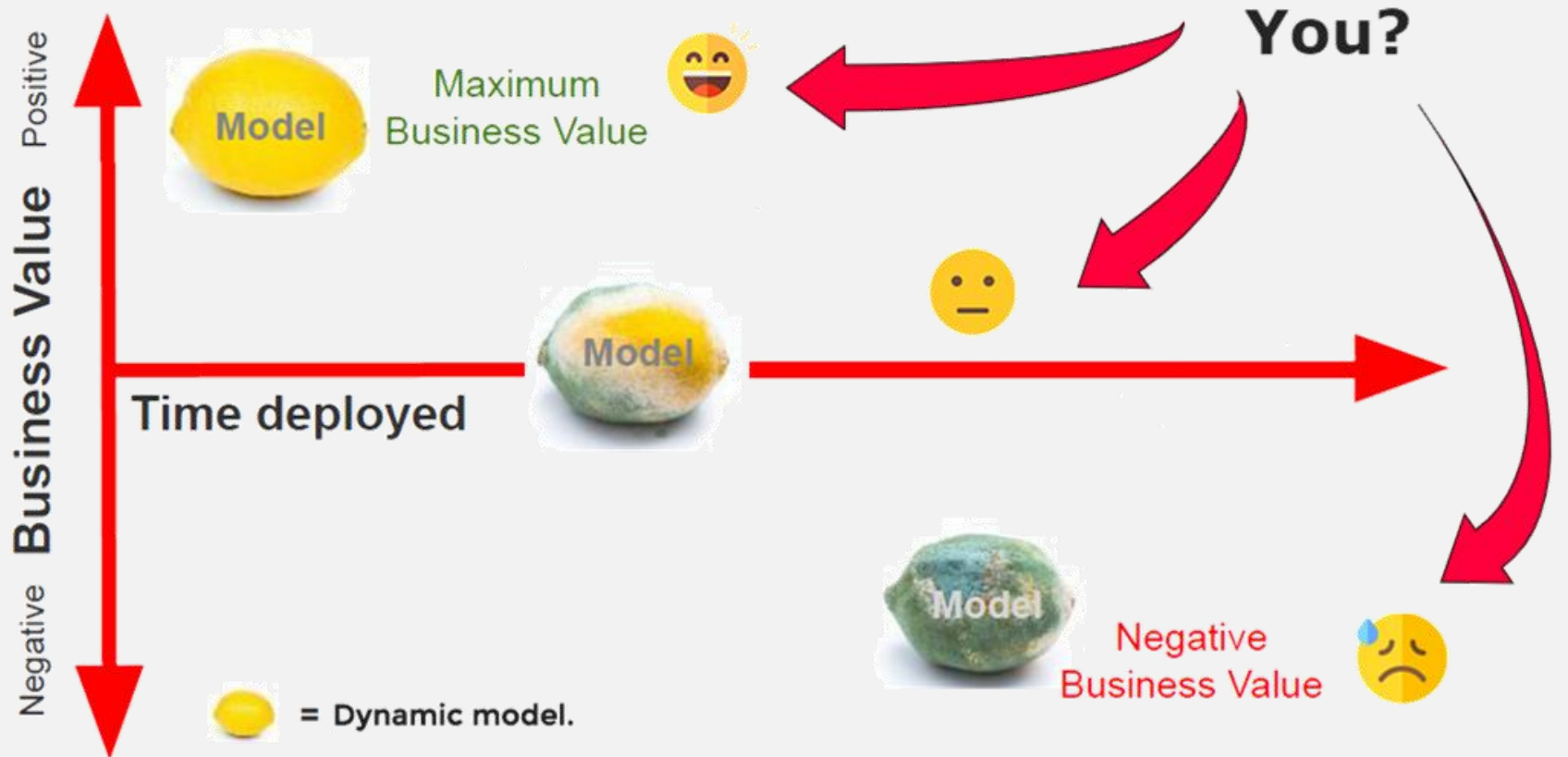
Monitoring

#Monitoring continu les performances des modèles de Machine Learning en production

UP ASI
Département informatique
Bureau: E204



Introduction



Small text: Copyright © 2019

Les modèles Machine Learning se dégradent au fil du temps. Ils sont dynamiques et sensibles aux changements réels du monde réel.

Machine Learning (sans monitoring)

- **Concept Drift (Dérive conceptuelle)** : Les relations entre les variables d'entrée et la cible peuvent évoluer au fil du temps, rendant les prédictions du modèle moins précises.
- **Data drift (Dérive des données)**: Les données réelles évoluent, ce qui peut rendre les modèles obsolètes ou moins pertinents.
- **Décisions non optimisées** : Sans suivi, les décisions basées sur des modèles dégradés peuvent nuire à la performance de l'entreprise.

Machine Learning (sans monitoring)

- **Solution:**



Monitoring : Définition

- **Monitoring des modèles en ML** : Pratique essentielle pour suivre les performances des modèles en production.
- **Objectif générale** : Garantir l'exactitude et la pertinence des modèles ML.
- **Importance** : Les modèles ML pilotent des décisions commerciales cruciales.

Monitoring : Objectifs

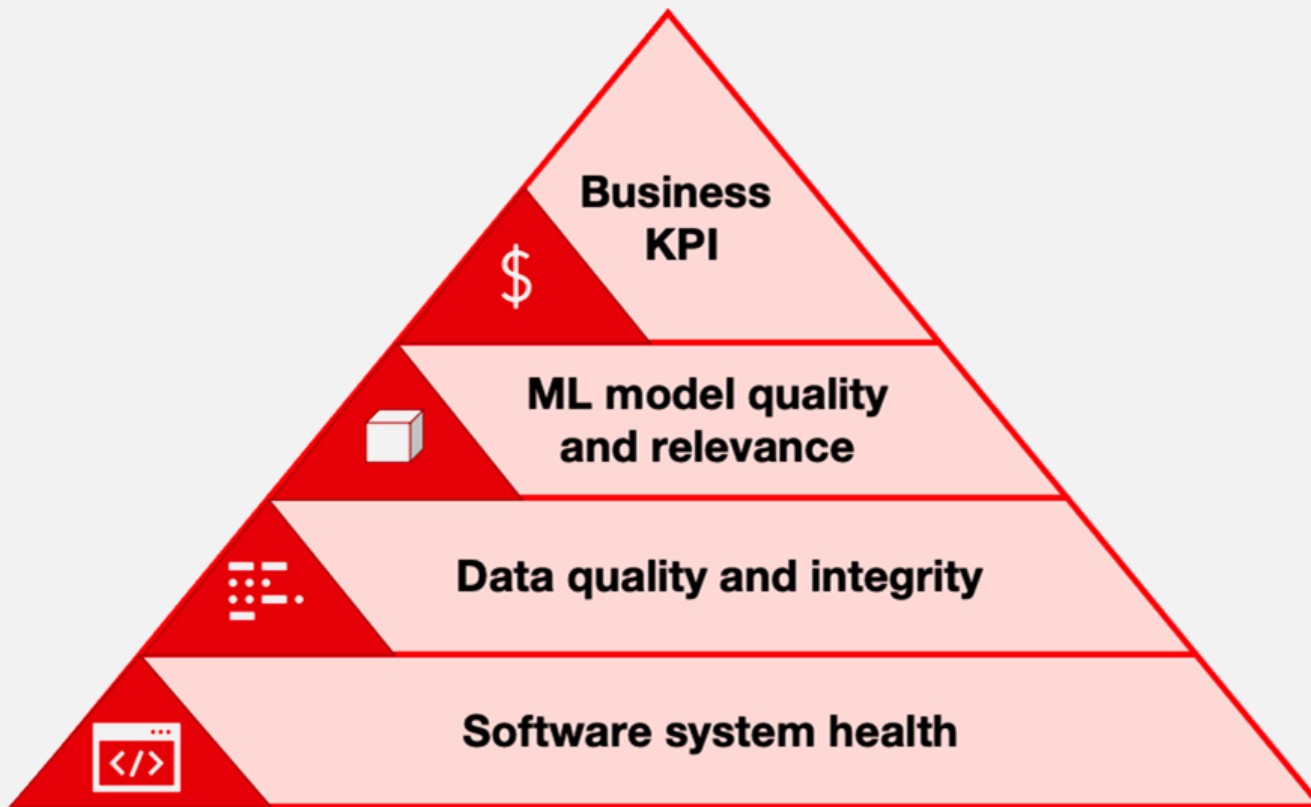
- Pour détecter les **problèmes** liés à un **modèle** et au système qui le sert **en production** avant qu'ils ne commencent à générer une valeur commerciale négative,
- Prendre des mesures en triant et en **dépannant les modèles en production** ou les intrants et systèmes qui les permettent,
- Pour garantir que leurs prévisions et leurs résultats puissent **être expliqués et rapportés**,
- Pour garantir que le processus de prédiction du modèle soit **transparent** pour les parties prenantes concernées pour une gouvernance appropriée,
- Pour fournir un chemin pour **maintenir** et **améliorer** le modèle en production.

Monitoring : Objectifs

Une application Machine Learning n'est pas seulement le modèle, mais tout ce qui permet au **modèle d'être en production**, y compris l'infrastructure, les données d'entrée, les ressources et d'autres services en amont et/ou en aval.

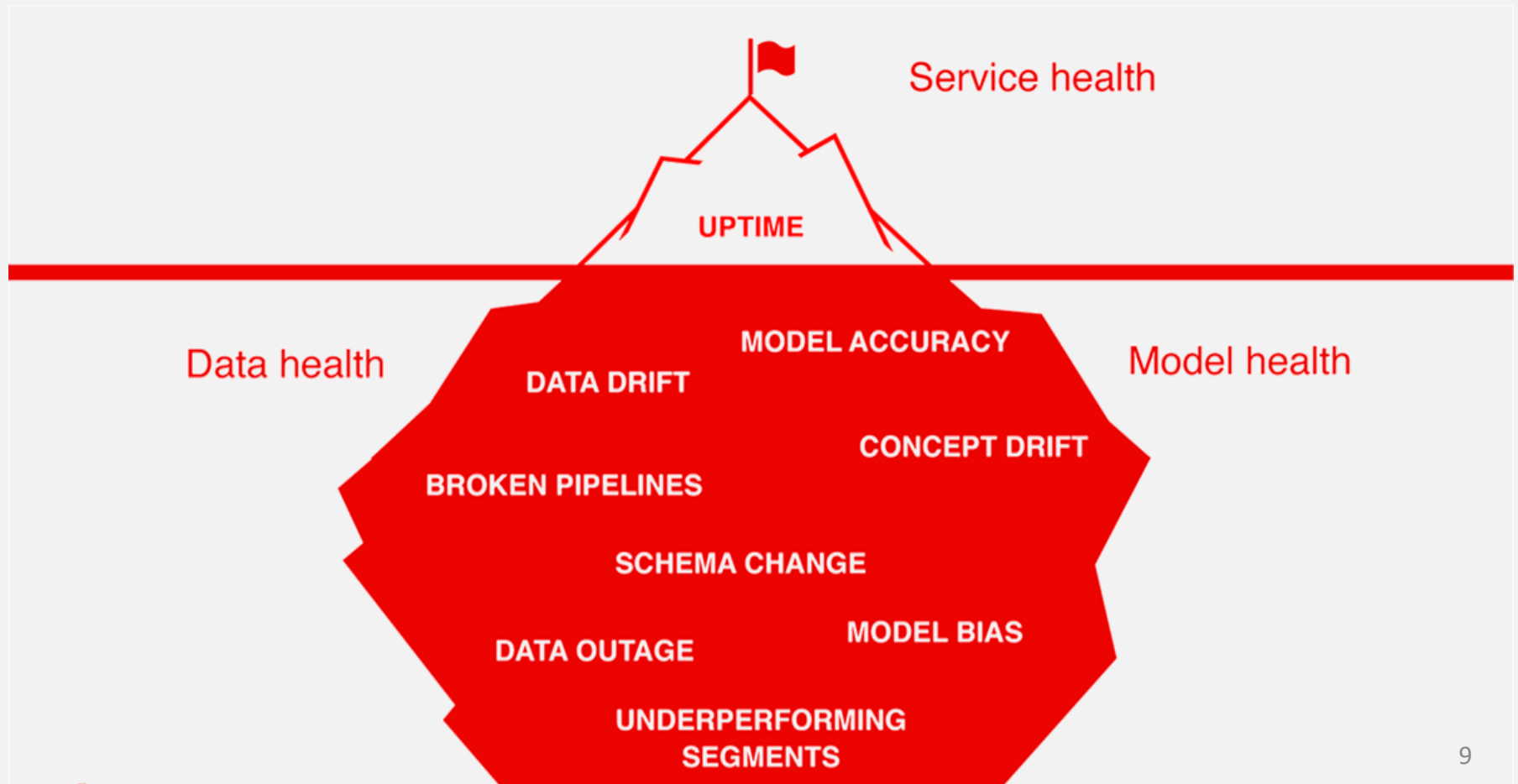
Composants de monitoring du système ML

- Les éléments à surveiller:



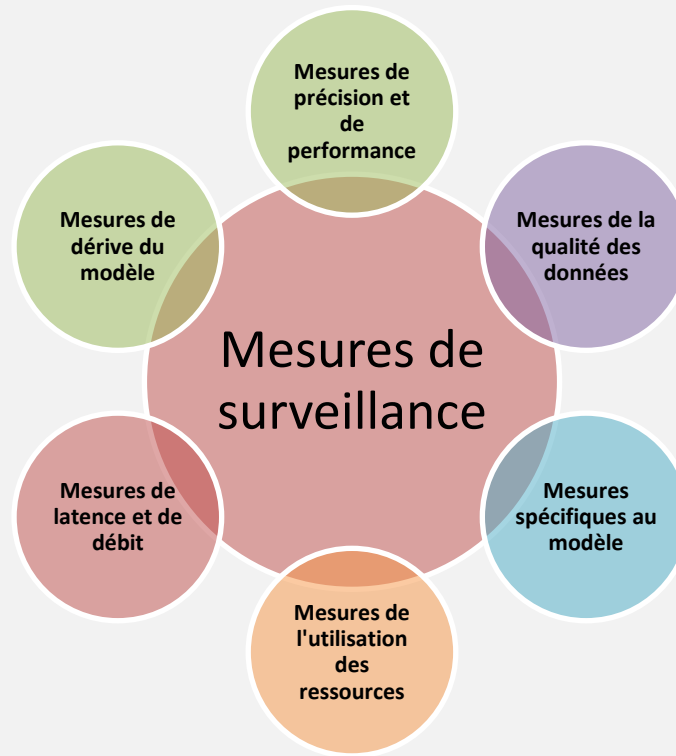
Composants de monitoring du système ML

- Les éléments à surveiller:



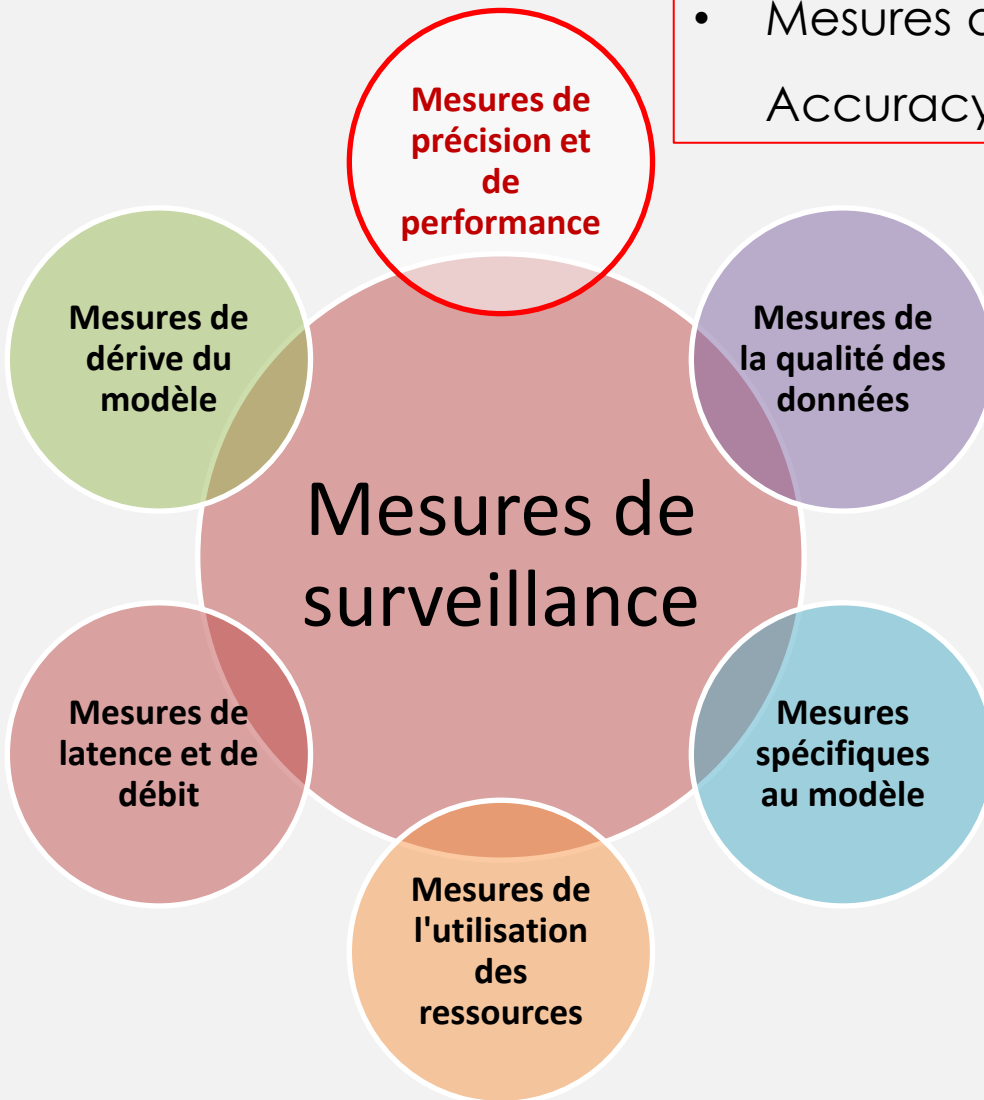
Mesures de surveillance MLOps

- La surveillance MLOps consiste à suivre diverses mesures pour évaluer les performances et le comportement des modèles ML.
- Ces mesures fournissent des informations précieuses sur la santé et l'efficacité des modèles en production.

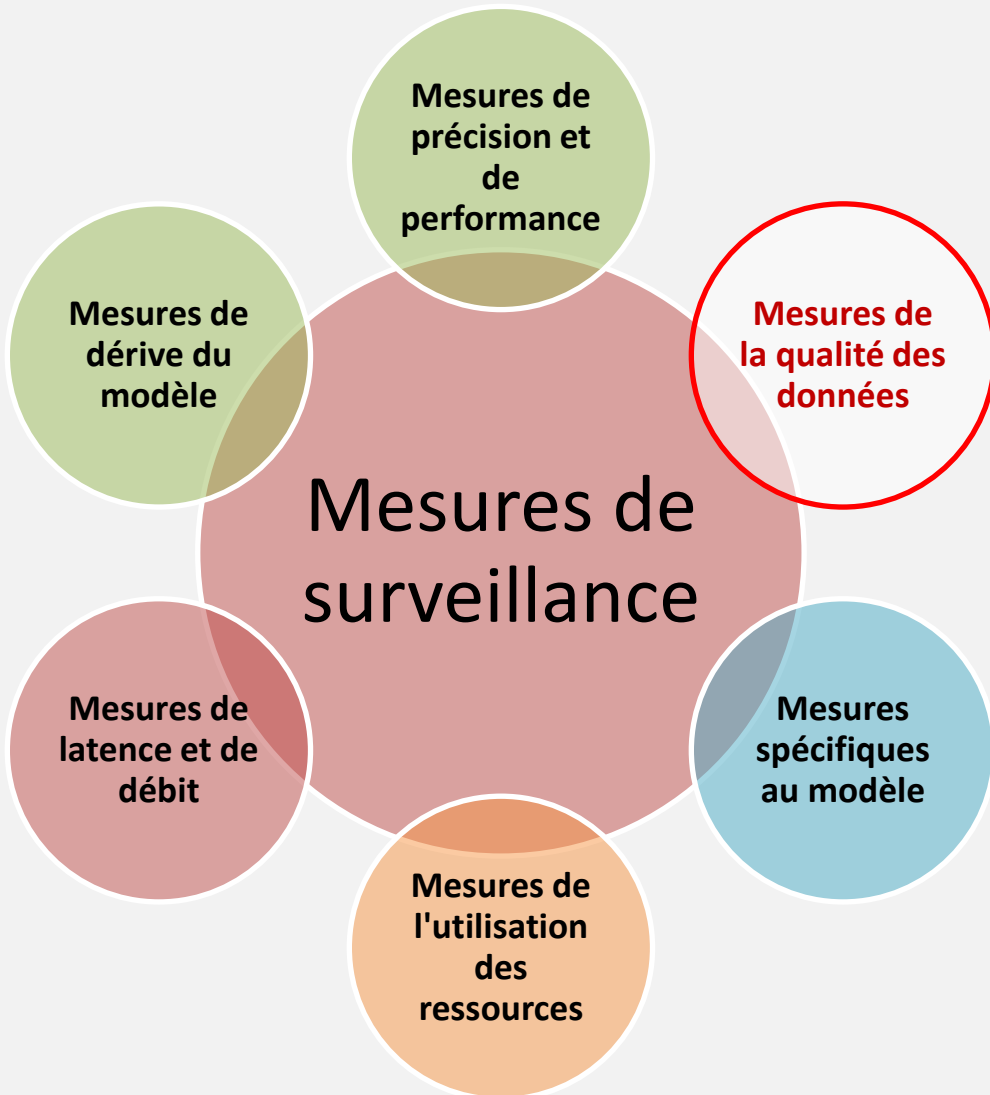


Mesures de surveillance MLOps

- Mesures de **performance** couramment utilisées: Accuracy, précision, recall, f1_score...

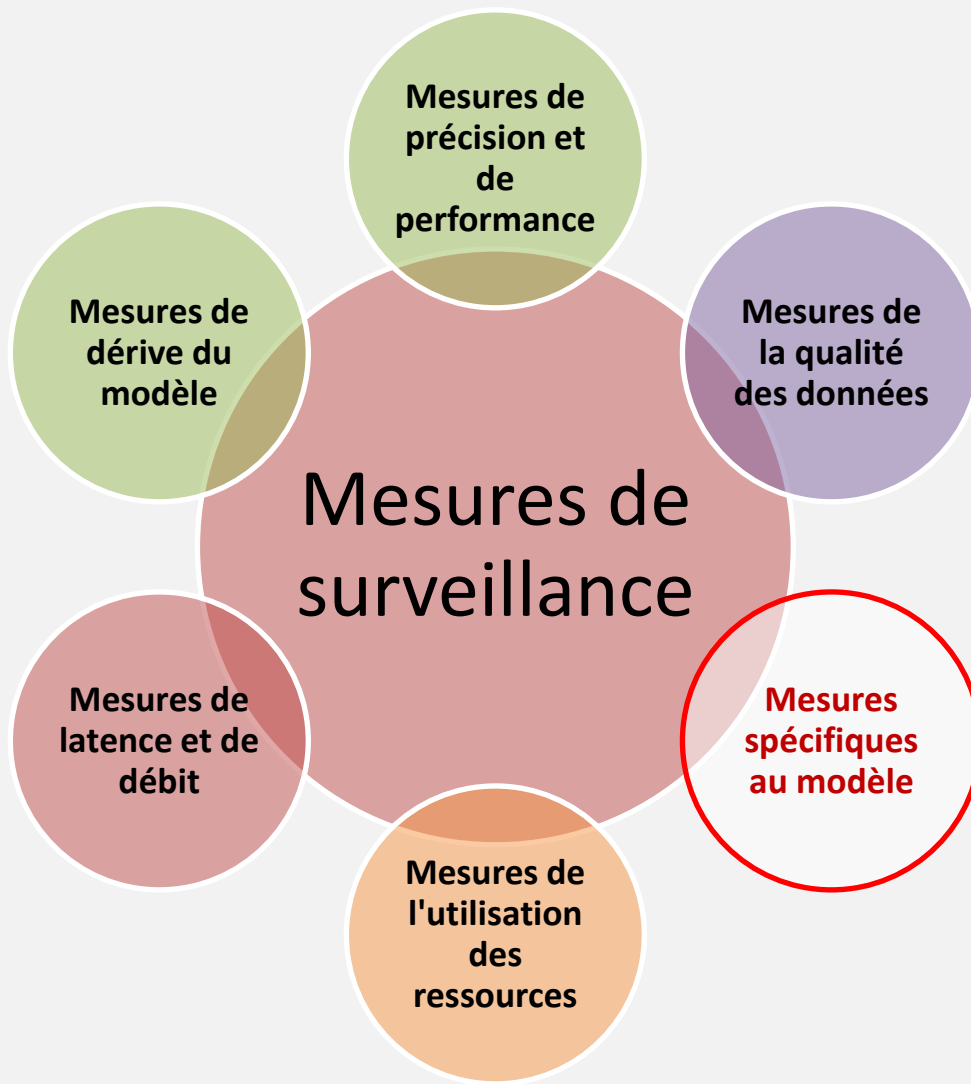


Mesures de surveillance MLOps



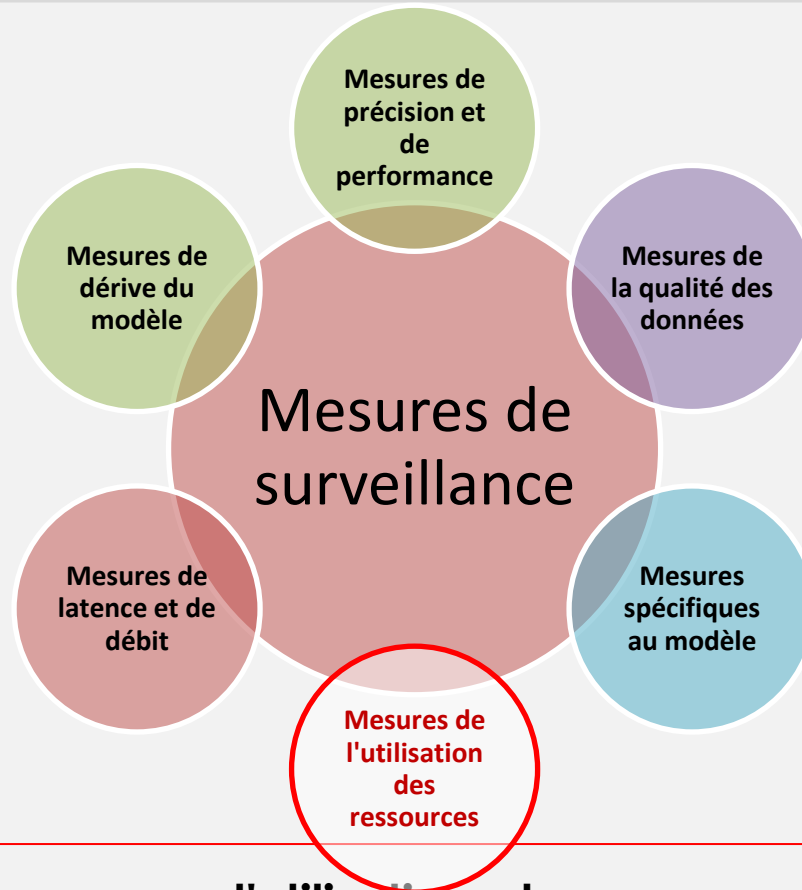
- Les **mesures de qualité des données** évaluent l'exhaustivité, la cohérence et la distribution des données pour identifier des problèmes comme les valeurs manquantes ou aberrantes.
- Leur surveillance est cruciale pour préserver l'intégrité des données et éviter les biais ou la dégradation des performances des modèles.

Mesures de surveillance MLOps



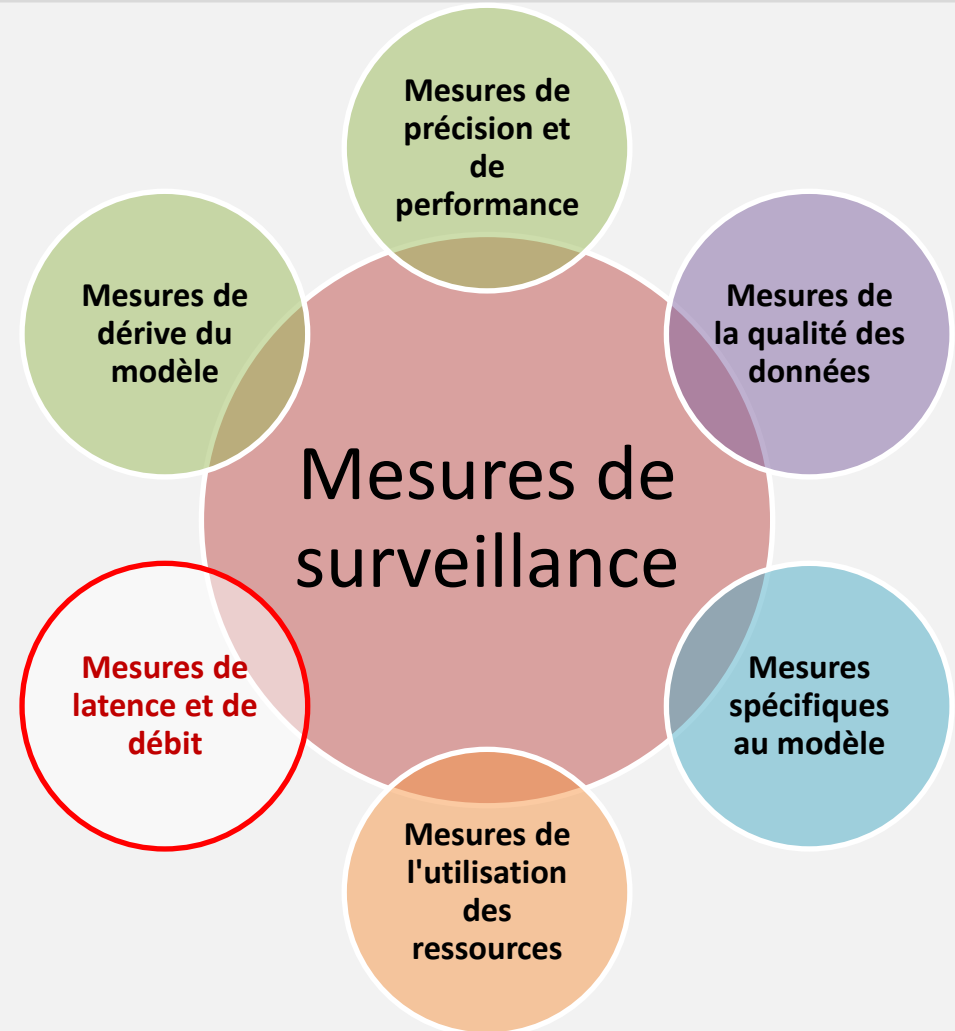
- Selon le type de modèle ML et le domaine du problème, il peut y avoir des **mesures spécifiques** qu'il est pertinent de surveiller.
- Exemple: BLEU et perplexité en NLP, la précision moyenne (mAP) et/ou l'intersection sur l'union (IoU) en computer vision.

Mesures de surveillance MLOps



- La surveillance **des mesures d'utilisation des ressources** permet d'évaluer l'efficacité et la rentabilité des modèles ML. Ces mesures incluent l'utilisation du processeur et du processeur graphique, la consommation de mémoire et les E/S de disque.

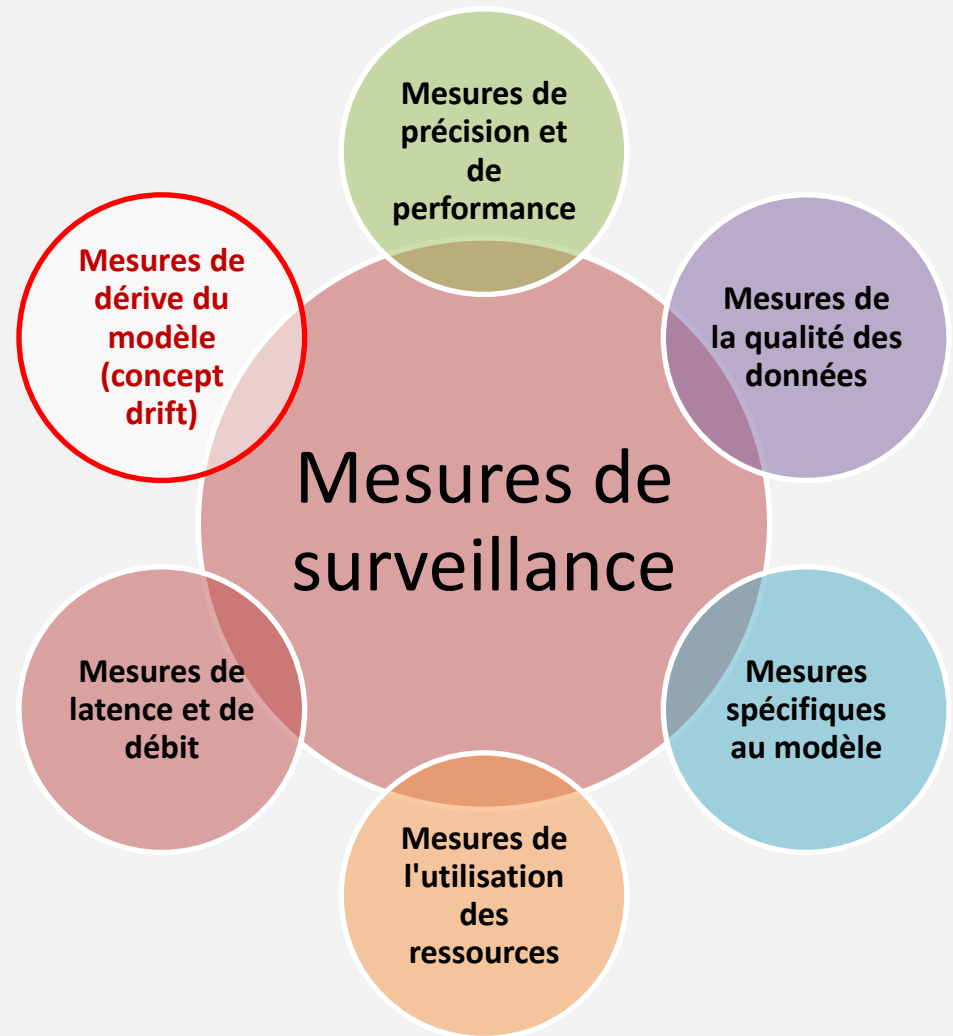
Mesures de surveillance MLOps



- Les **mesures de latence et de débit** mesurent le temps de réponse et la capacité de traitement des modèles ML.

Mesures de surveillance MLOps

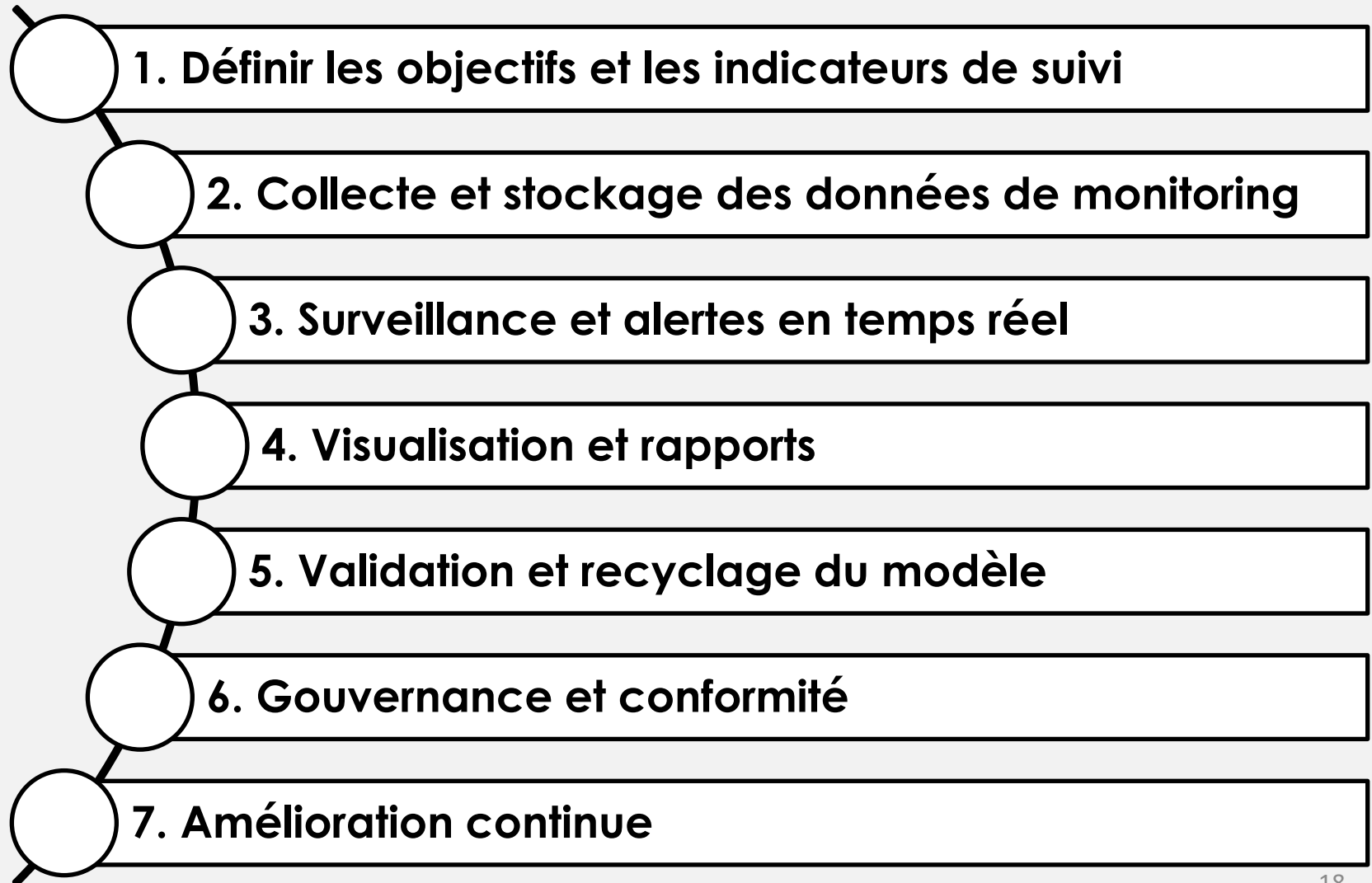
- Les **mesures de dérive de modèle** permettent de détecter les changements de comportement ou de performance des modèles ML au fil du temps.
- Ces mesures mesurent l'écart entre les prédictions ou les résultats du modèle et le comportement de base attendu.



Framework de monitoring du modèle MLOps

- Pour mettre en œuvre une surveillance MLOps efficace, les organisations peuvent établir un **Framework de surveillance** des modèles qui décrit les processus, les outils et les responsabilités de surveillance des modèles ML en production.
- Les composants d'un Framework de surveillance sont définis selon l'organisation.

Les composants d'un Framework de monitoring du modèle MLOps



Quelques outils de monitoring continu

Catégorie	Outils
Suivi des expériences et gestion des modèles	MLflow Tracking , Weights & Biases (W&B), Neptune.ai
Monitoring des performances des modèles	Evidently AI, WhyLabs (WhyLogs), Fiddler AI, Arize AI
Logs et monitoring des infrastructures	OpenSearch Dashboards, Prometheus + Grafana, Datadog AI Monitoring, Seldon Alibi Detect
Déploiement et gestion des modèles	KServe, Seldon Core, BentoML, Triton Inference Server
Suivi des logs et tracing	Opentracing / Jaeger, ELK Stack (Elasticsearch, Logstash, Kibana) , Fluentd / Fluent Bit
Automatisation et orchestration	Kubeflow Pipelines, MLflow Model Registry , Metaflow

To do

- Effectuer le monitoring en utilisant **Elasticsearch** et **KIBANA** :
 1. Installation d'Elasticsearch et Kibana avec Docker
 2. Envoyer des logs de modèle à Elasticsearch
 3. Visualiser les logs dans Kibana

Monitoring

Si vous avez des questions, n'hésitez pas à nous contacter :

Département Informatique

UP Architectures des Systèmes d'Information

Bureau E204 /E304