



EXPLICABILITÉ DE L'IA EN FINANCE

-

LE POINT DE VUE DU SUPERVISEUR

Laurent Dupont
Pôle Fintech-Innovation, ACPR

-
9 novembre 2020



Objectifs

1. Promouvoir un développement maîtrisé de l'IA en finance
2. Préparer les superviseurs au contrôle de l'IA



Décembre 2018

Intelligence artificielle : enjeux pour le secteur financier

Document de réflexion

AUTEURS
Olivier FLICHE, Su YANG - Pôle Fintech-Innovation, ACPR



Juin 2020

Gouvernance des algorithmes d'intelligence artificielle dans le secteur financier

Document de réflexion

AUTEURS
Laurent DUPONT, Olivier FLICHE, Su YANG
Pôle Fintech-Innovation, ACPR



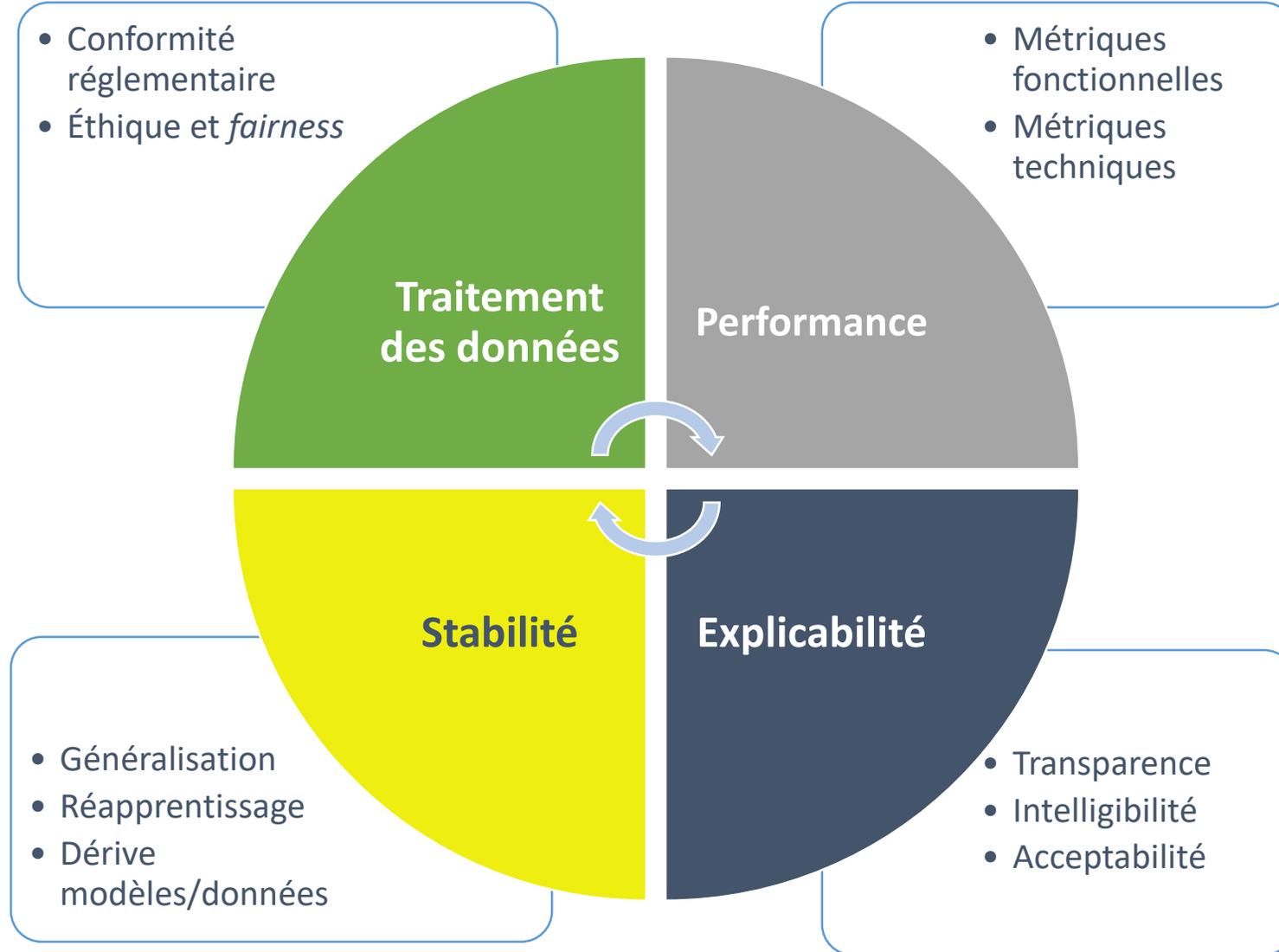


PRINCIPES TECHNIQUES DE CONCEPTION DE L'IA

PÔLE FINTECH-INNOVATION

9 NOVEMBRE 2020

LES 4 PRINCIPES DE CONCEPTION DE L'IA



PRINCIPES DE CONCEPTION DE L'IA

- Conformité réglementaire
- Éthique et *fairness*

Traitement
des données

“ *We will tackle unconscious bias that exists in people, institutions and even in algorithms.*

Ursula von der Leyen
(State of the Union, 16 Sep. 2020)

“ *Algorithms must not be a black box and there must be clear rules if something goes wrong.*

Ursula von der Leyen
(State of the Union, 16 Sep. 2020)

Explicabilité

- Transparence
- Intelligibilité
- Acceptabilité

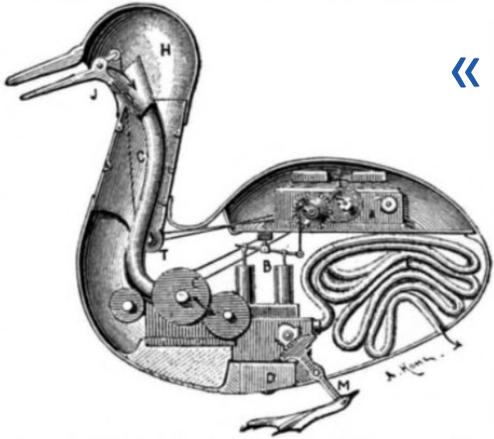


EXPLICABILITÉ DE L'IA EN FINANCE

PÔLE FINTECH-INNOVATION

9 NOVEMBRE 2020

LES DIMENSIONS DE L'EXPLICABILITÉ

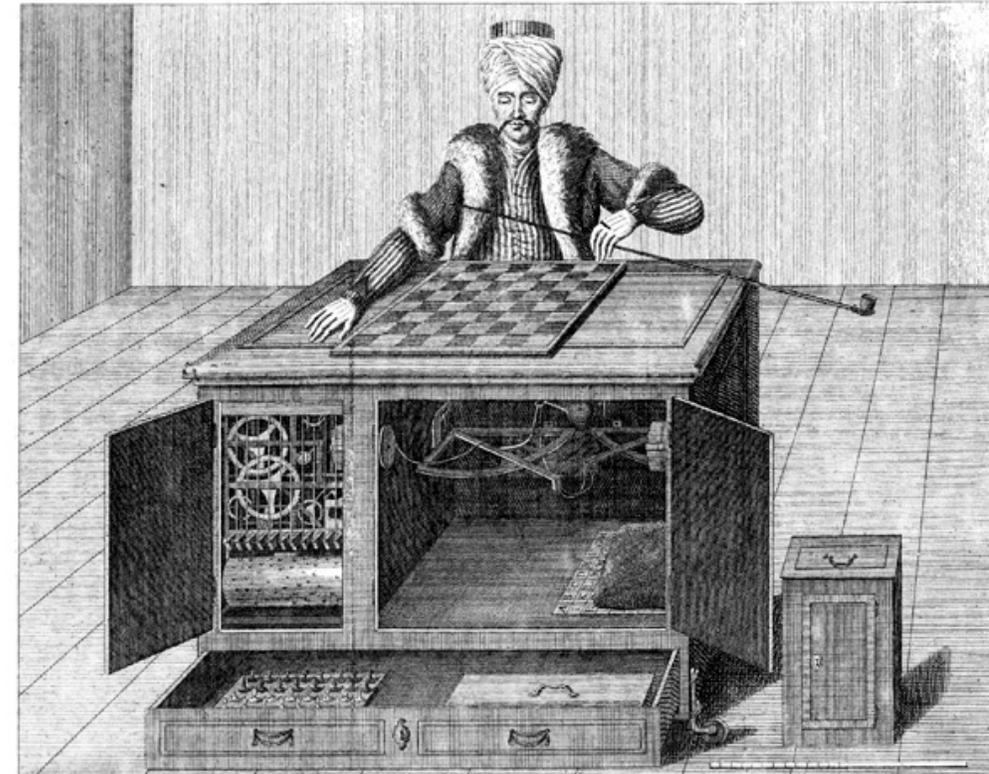


« Comment ? » : la transparence

- Validation et surveillance internes
- Auditabilité par le superviseur

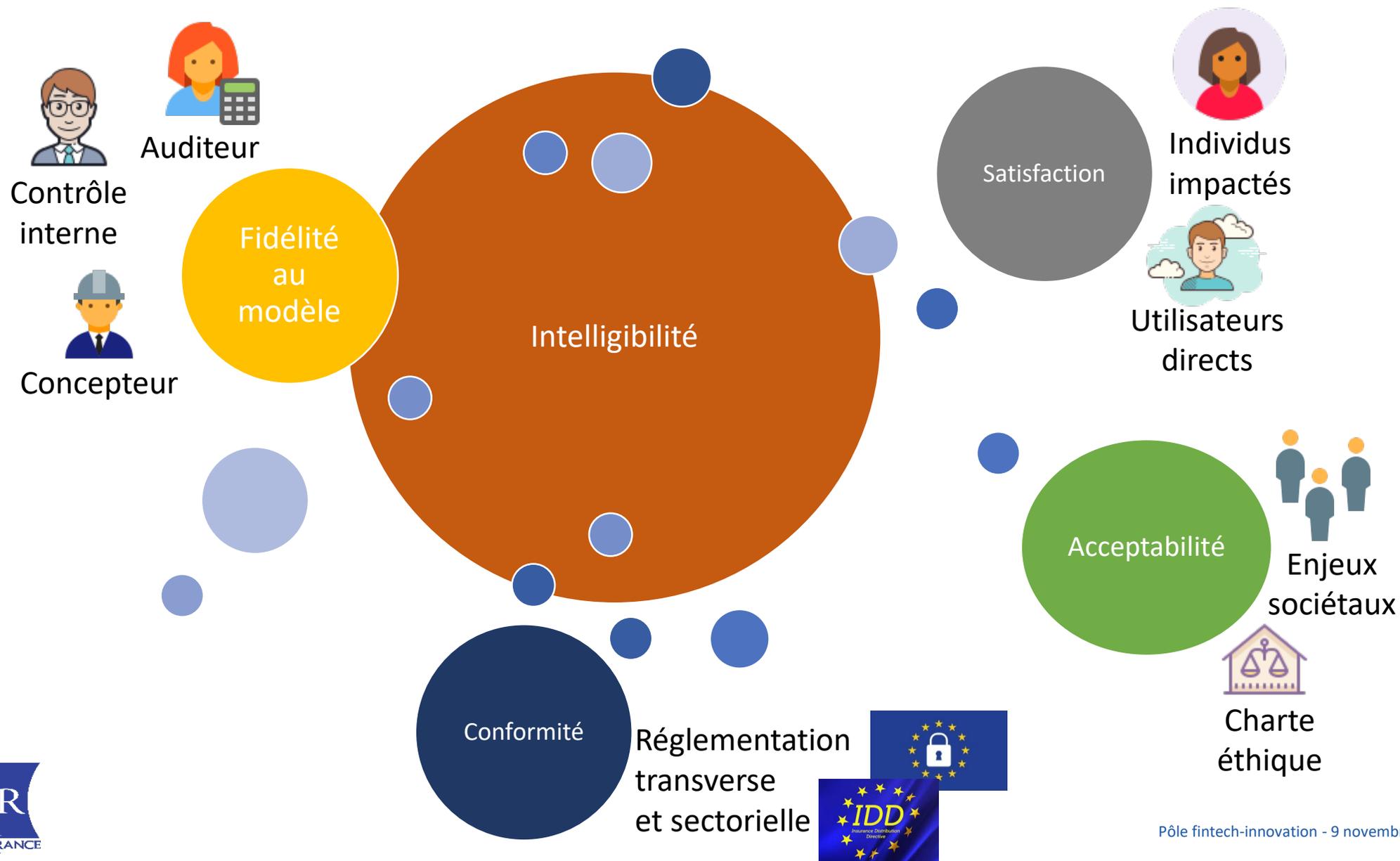
« Pourquoi ? » : l'interprétabilité

- Compréhension du comportement du système par les opérateurs humains qui interagissent avec lui
- Compréhension par le client ou utilisateur final
- Acceptabilité sociale ou éthique de la solution

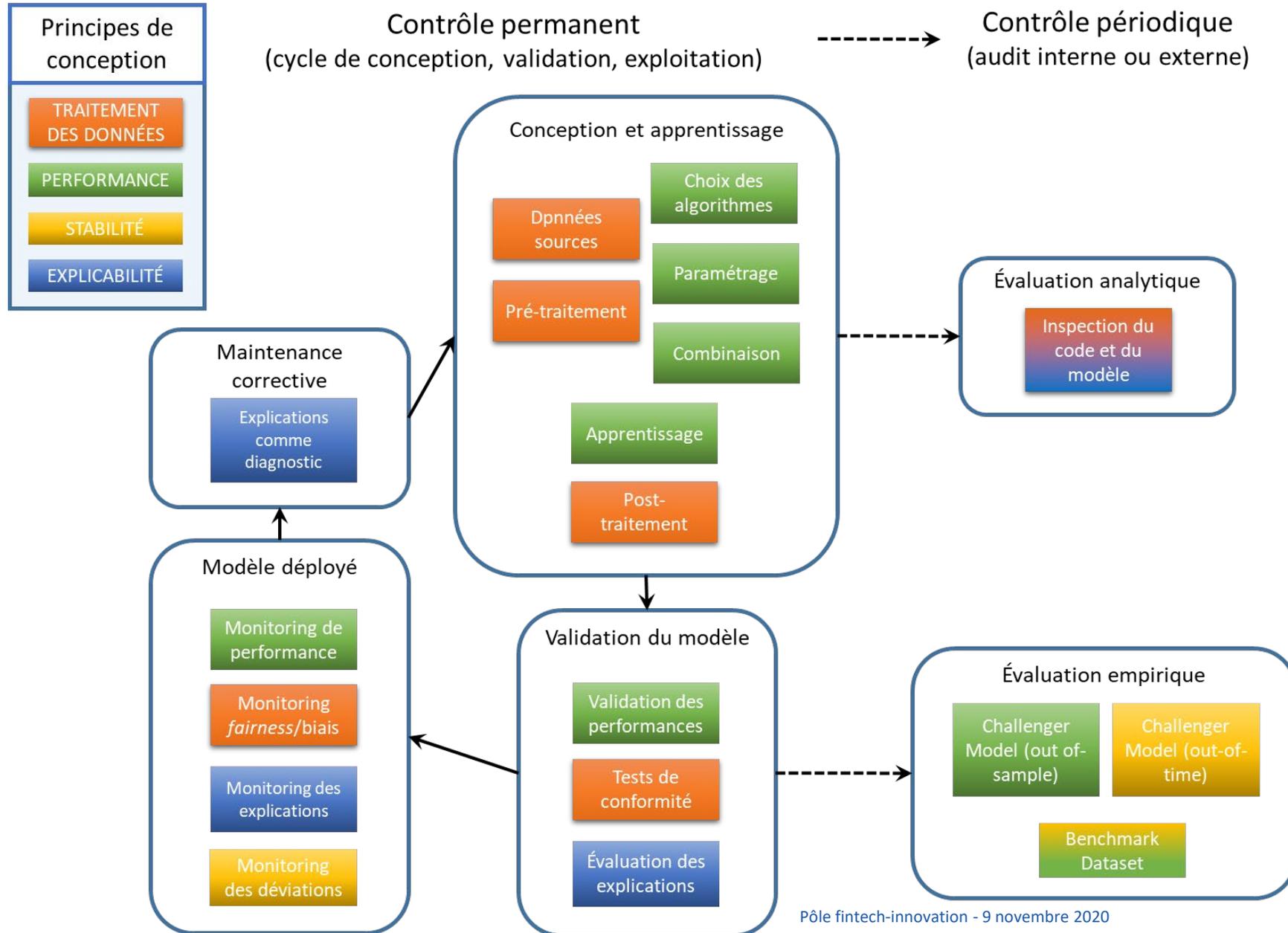




TENSIONS ENTRE OBJECTIFS D'EXPLICABILITÉ



ÉVALUATION DU ML AU COURS DE SON CYCLE DE VIE





ÉCHELLE DE NIVEAUX D'EXPLICATION

Explication de niveau 1 : observation

Elle répond sous un angle technique à la question : « *Que fait l'algorithme ?* », ou sous un angle plus fonctionnel : « *À quoi sert l'algorithme ?* ». Ce niveau d'explication peut être obtenu :

- de
- (in
- de
- de
- un

Explication de niveau 2 : justification

Elle répond à la question : « *Pourquoi l'algorithme donne-t-il tel résultat (en général ou dans une situation précise) ?* ». Ce niveau d'explication peut être obtenu :

- soit
- 4), é
- cont
- soit
- appr

Explication de niveau 3 : approximation

Elle fournit une réponse, souvent inductive, à la question : « *Comment fonctionne l'algorithme ?* ». Ce niveau d'explication peut être obtenu, en sus des méthodes des niveaux 1 et 2 :

- pa
- M
- pa
- se
- de
- hy

Explication de niveau 4 : réplication

Elle fournit une réponse démontrable à la question : « *Comment prouver que l'algorithme fonctionne correctement ?* ».

Ce niveau d'explication peut être obtenu, en sus des méthodes des niveaux 1 à 3, par une analyse détaillée de l'algorithme, des modèles et des données. Dans la pratique, cela n'est possible que par une revue ligne à ligne du code source, une étude exhaustive des jeux de données utilisés, et un examen de l'ensemble des paramètres du modèle.



EXEMPLE DE NIVEAU D'EXPLICATION

Audience de l'explication	Contexte	Risque associé	Niveau d'explication requis
Client	Processus d'indemnisation	Risque opérationnel (insatisfaction du client)	1
Contrôleur interne	Vérification au quotidien du bon fonctionnement du processus	- Risque opérationnel - Risque de conformité (respect du contrat) - Risque financier	2
Auditeur	Évaluation de l'algorithme	- Risque opérationnel - Risque de conformité (respect du contrat) - Risque financier	3

Domaine

Contrats d'assurance

Processus métier

Gestion de contrat

Fonctionnalité

Propositions d'indemnisation



MÉTHODOLOGIE DE L'IA EXPLICABLE

PÔLE FINTECH-INNOVATION

9 NOVEMBRE 2020



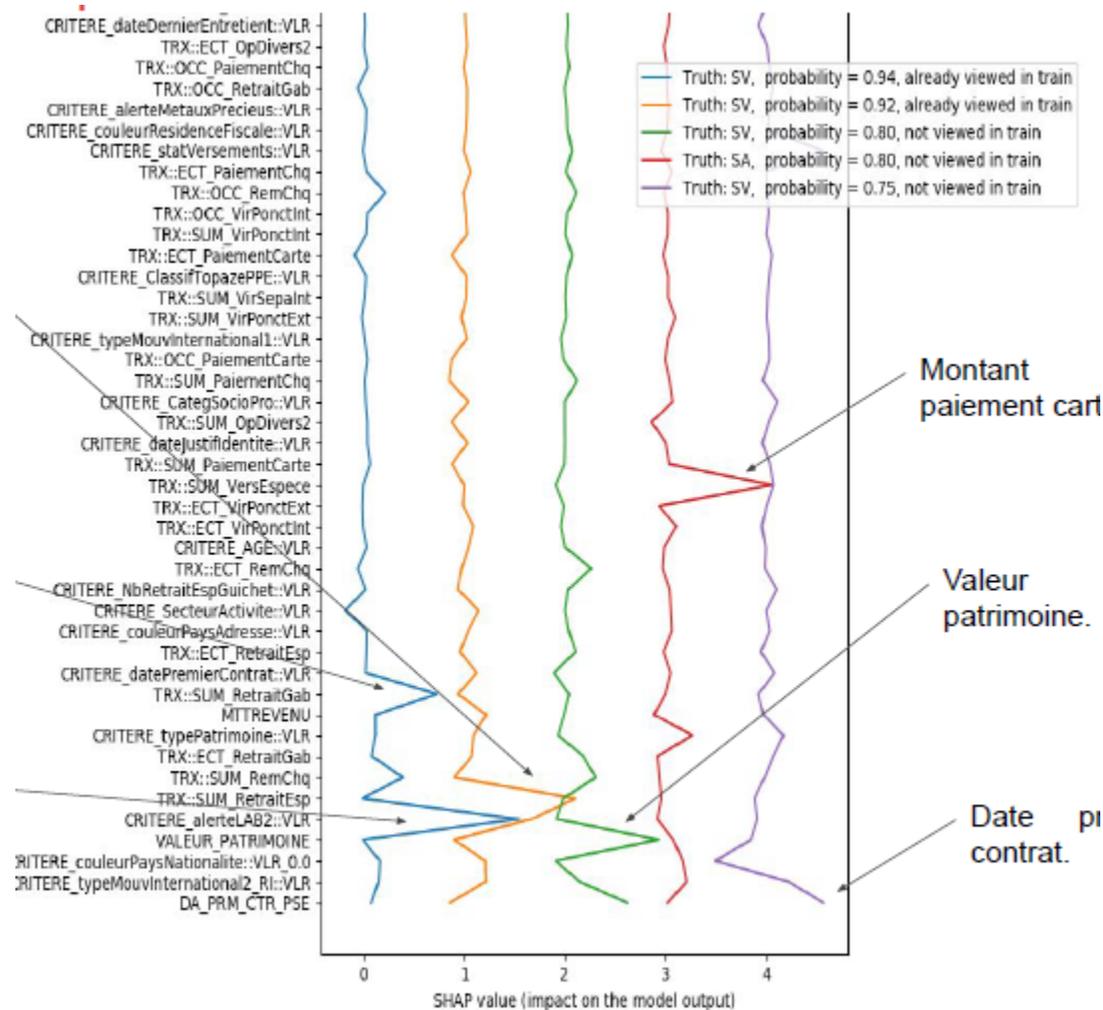
EXPLICATIONS POST-MODÉLISATION

• Méthodes les plus courantes

- *Partial Dependency Plots*
- LIME
- SHAP et variantes
- *Global Surrogate Models*

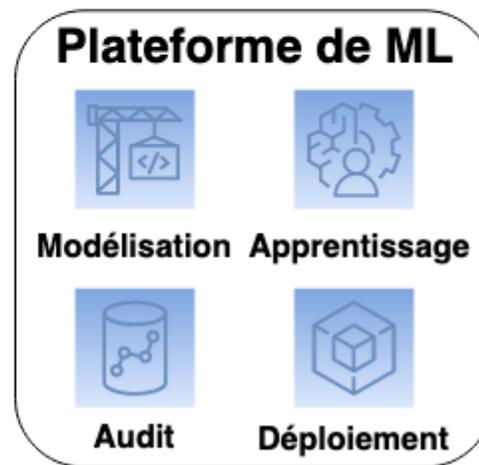
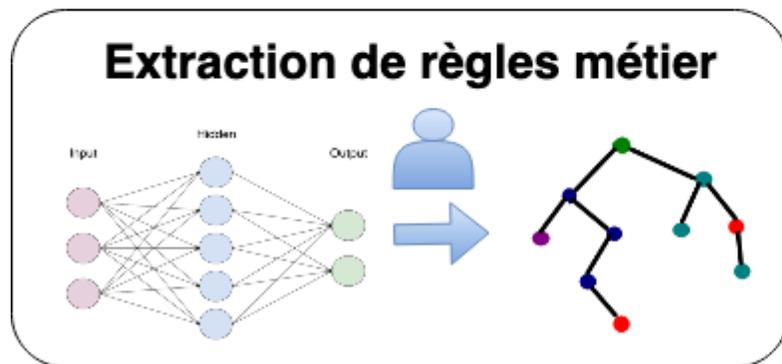
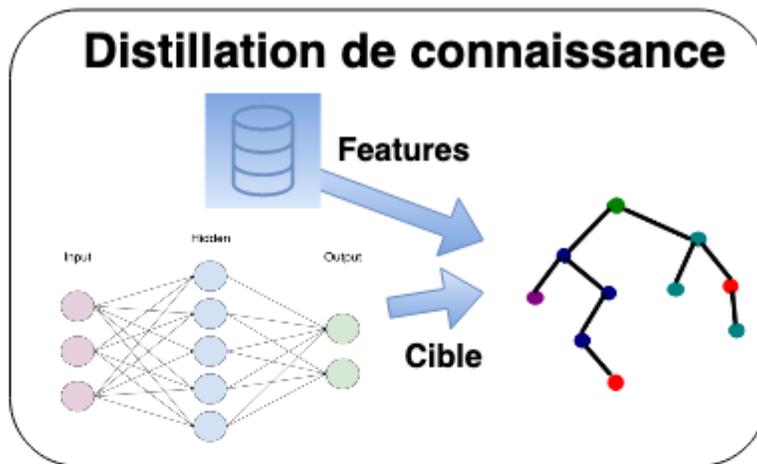
• Avant-garde

- Anchors
- MACEM (contrefactuelles)



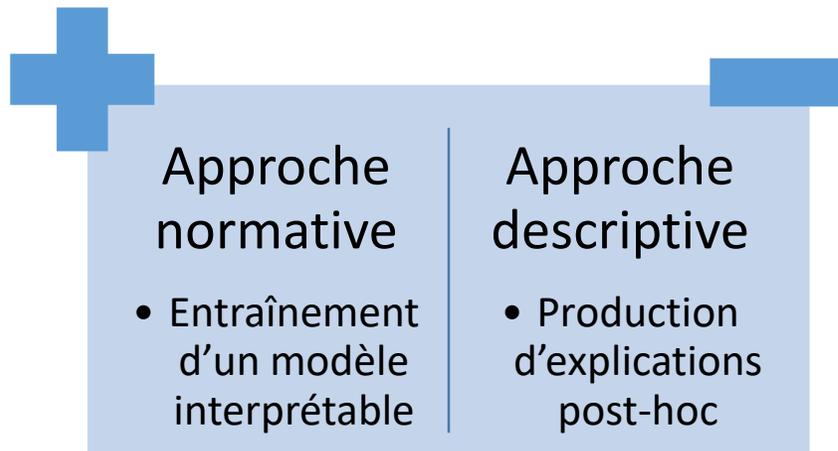


ARCHITECTURES EXPLICABLES



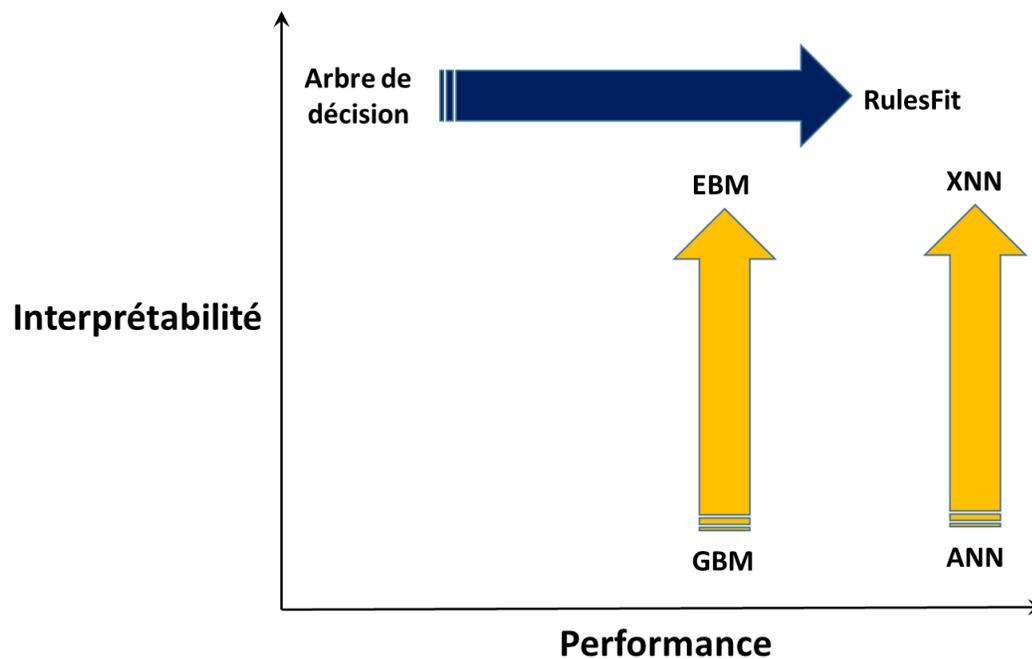


MODÈLES INTERPRÉTABLES



“Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.”

Cynthia Rudin (2018)





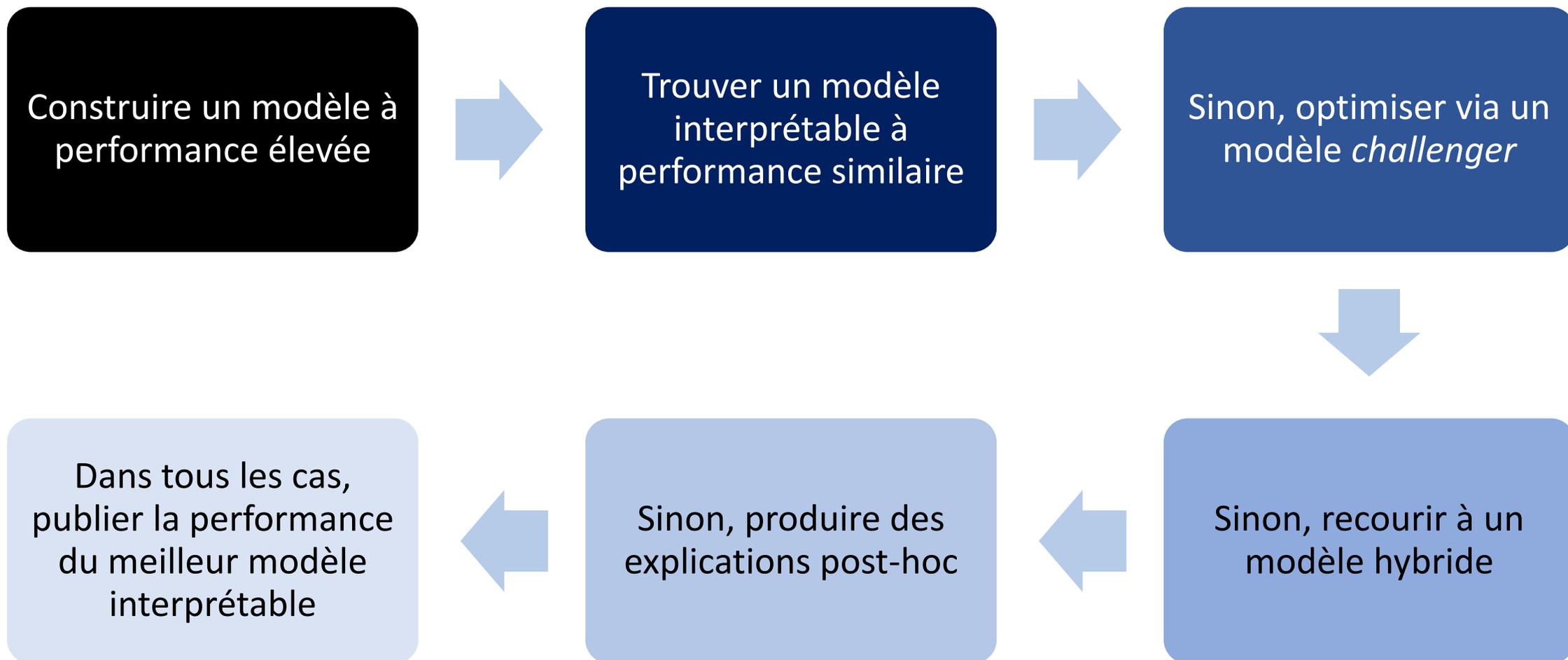
SUPERVISION DE L'IA EN FINANCE

PÔLE FINTECH-INNOVATION

9 NOVEMBRE 2020



APPROCHE PRÉCONISÉE POUR L'IA EXPLICABLE





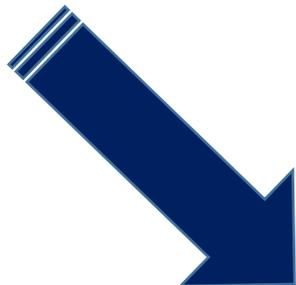
LIMITES À L'EXPLICABILITÉ (*THE RETURN OF THE BLACKBOX*)



Apprentissage non supervisé

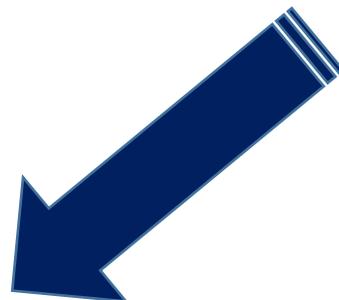


NLP

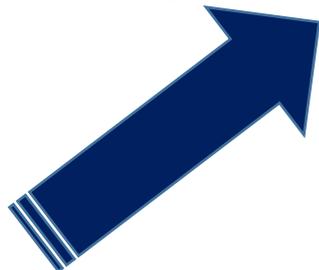


Algorithms that remember: model inversion attacks and data protection law.

Michael Veale et al. (2018)



Vendor models



NIST Study Shows Face Recognition Experts Perform Better With AI as Partner



PROPOSITION DE « BOÎTE À OUTILS DU SUPERVISEUR D'IA »

Objet audité

Approches analytiques



Approches empiriques

Outil d'audit...	...des données	...de l'algorithme ou service d'IA	...du modèle
<i>Standards descriptifs</i>	<i>Datasheets, Data Statements, etc.</i>	<i>Factsheets</i>	<i>Model Cards</i>
<i>Analyse statique</i>	Analyse exploratoire	Code source	Quelques standards en <i>Deep Learning</i> uniquement
<i>Méthodes explicatives</i>	Pré-modélisation	Méthodes conjointes (TED, etc.)	Post-modélisation (locales/globales, spécifiques/agnostiques au modèle)
<i>Méthodes empiriques d'audit</i>	Jeux de données de <i>benchmarking</i>	Algorithmes <i>challengers</i>	Modèles <i>challengers</i>



EXPLICABILITÉ DE L'IA EN FINANCE : *LE POINT DE VUE DU SUPERVISEUR*

Merci pour votre attention.

Liens utiles

- [Document de réflexion et consultation publique](#)
- [Version en anglais](#)
- [Chaire XAI4AML](#)

Pour toute question : **fintech-innovation@acpr.banque-france.fr**