WISG

Workshop interdisciplinaire sur la sécurité globale

14 & 15 mars 2024 • Couvent des Jacobins **Rennes**

EQUIHid : Vers une protection plus équitable de la vie privée pour les services de santé

Nesrine Kaaniche

Télécom SudParis

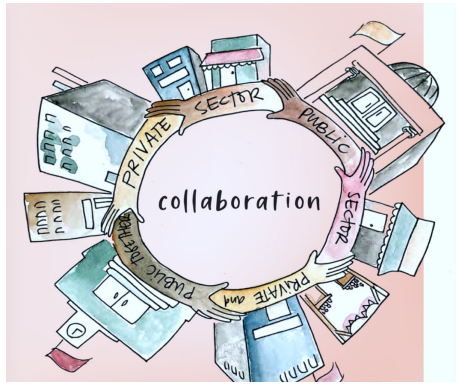
EQUIHid

• Un projet mono-partenaire – JCJC (Télécom SudParis, Institut Polytechnique de Paris) avec des collaborations:

•2 collaborations académiques:
- •LIST, CEA, Université Paris-Saclay
- •Université du Québéc, Montréal

• Un comité consultatif:
- •2 partenaires académiques/de recherche: Génépole d'Evry, Télécom Paris
- •2 partenaires industriels: Relyens, Medtronic

Les techniques de respect de vie privée renforcent-elles les inégalités sociales? Quel impact pour notre santé?



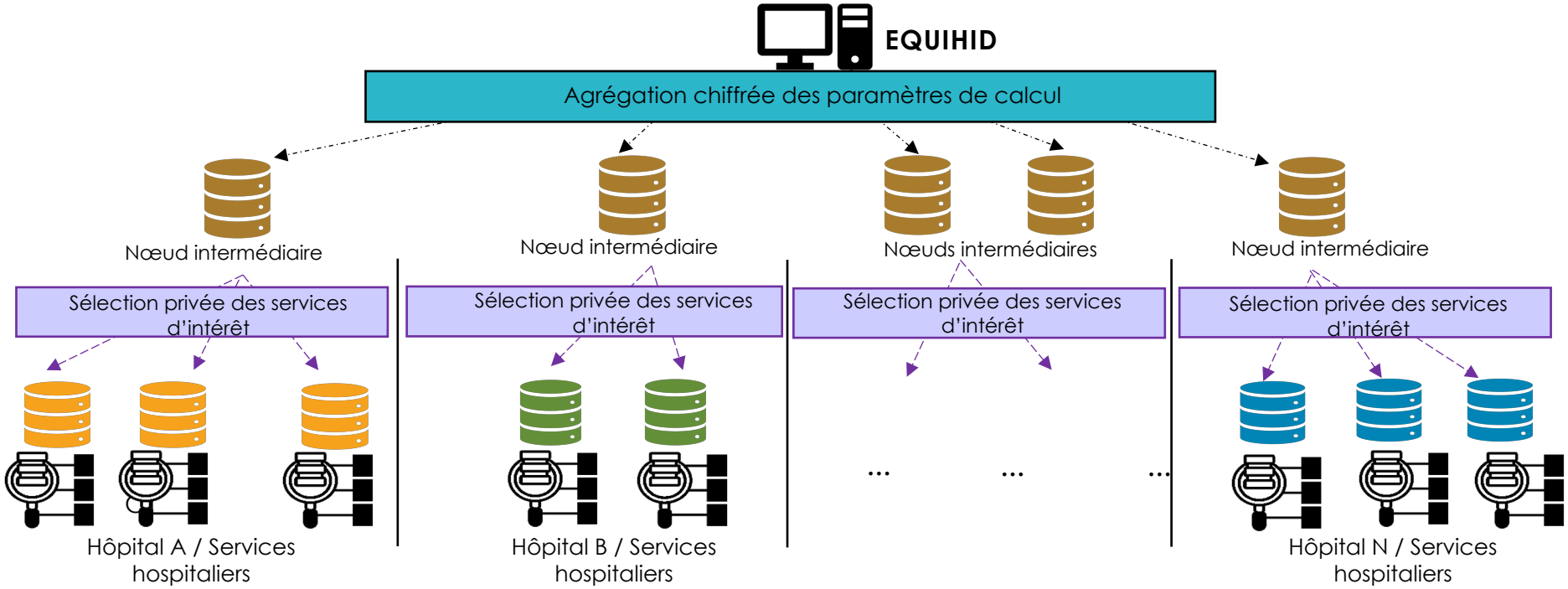
- **Des applications basées sur l'IA développées avec un souci limité de la vie privée des utilisateurs**
  - ➢ Violation massive de données sensibles
- **Difficile d'utiliser de larges bases de données, équilibrées et diverses, nécessaires pour garantir de meilleurs résultats des systèmes**

1. **Protection de la vie privée :**
   - L'apprentissage fédéré permet un calcul sur les données locales (chez le client)
   - Plusieurs attaques à la vie privée, notamment à partir des updates (Gradients, paramètres …):
     - ➢ Reconstruction des données, inférence d'informations sensibles, reconstitution des profils des patients.

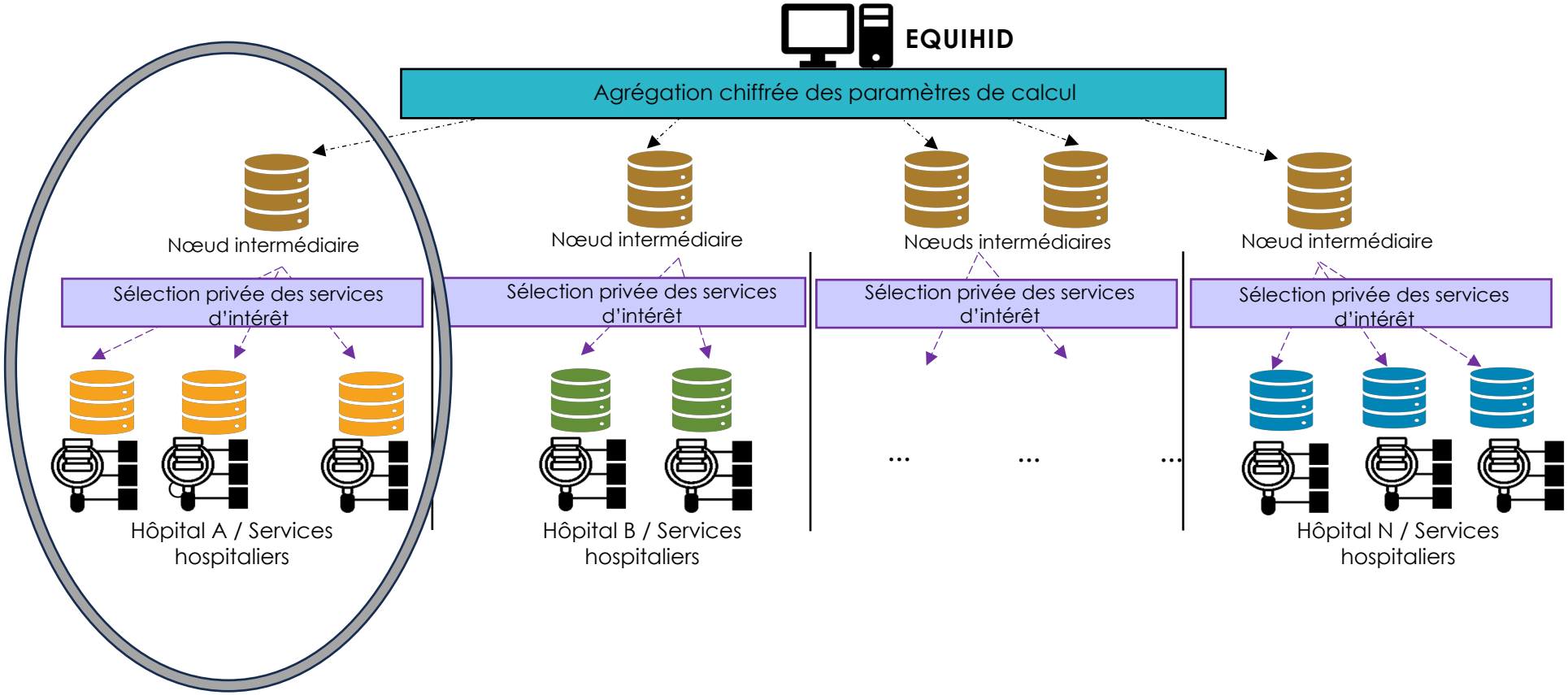
2. **Equité des modèles :**
   - Des algorithmes d'apprentissage développés sans considération majeure des problèmes d'équité
     - ➢ Des algorithmes augmentant des risques d'identification des classes minoritaires de 20%.

⇒ **Assurer un compromis entre équité, respect de la vie privée et utilité.**

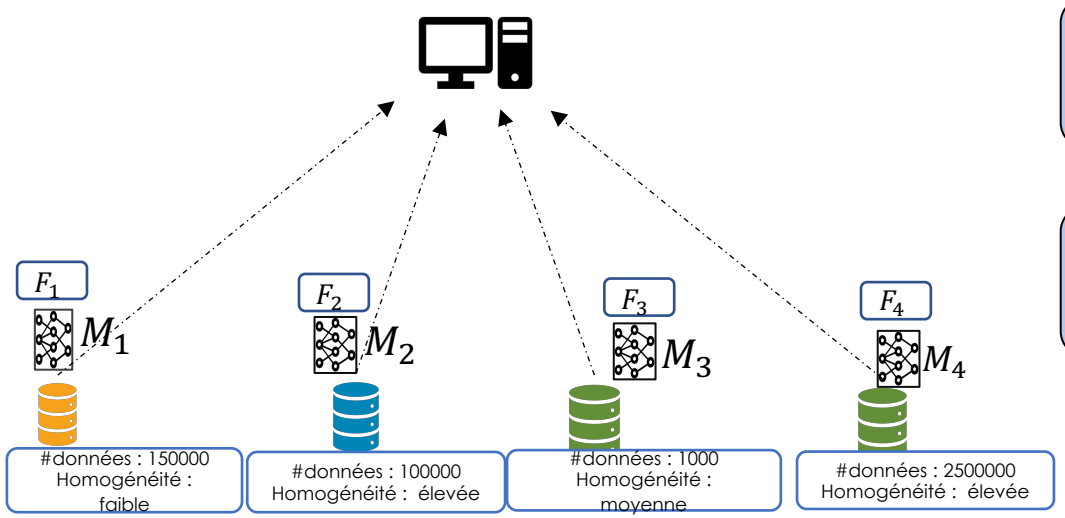
Architecture



Architecture



Approche et méthodologie



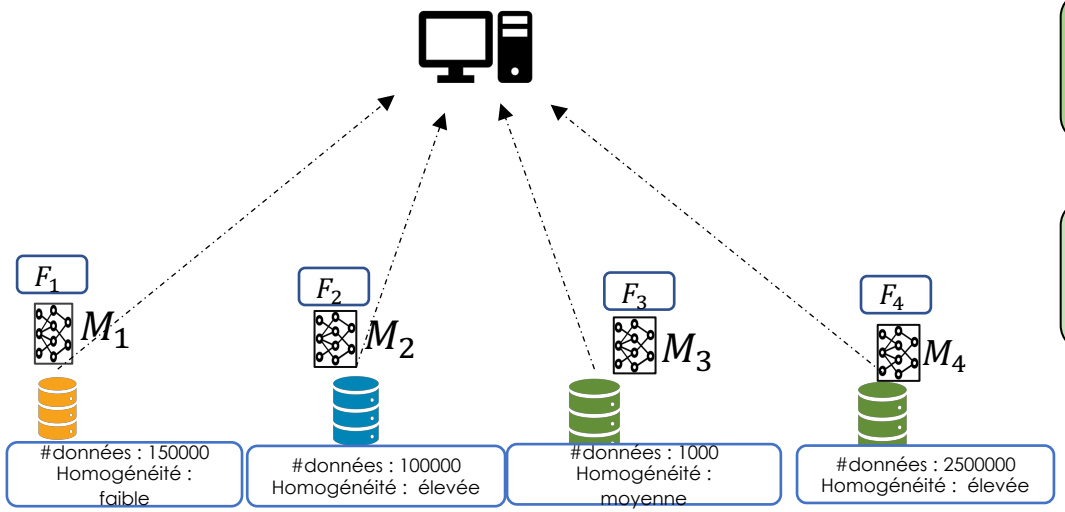
Cas d'usage santé : définition des métriques d'équité et besoins de sécurité.

Etude et proposition des attaques sur la confidentialité des modèles et des données, incluant les métriques d'équité.

Conception d'une approche d'agrégation équitable et respectueuse de la vie privée pour l'apprentissage fédéré

Implémentation et évaluation des solutions proposées par rapport aux attaques identifiées.

Approche et méthodologie



Cas d'usage santé : définition des métriques d'équité et besoins de sécurité.

Etude et proposition des attaques sur la confidentialité des modèles et des données, incluant les métriques d'équité.

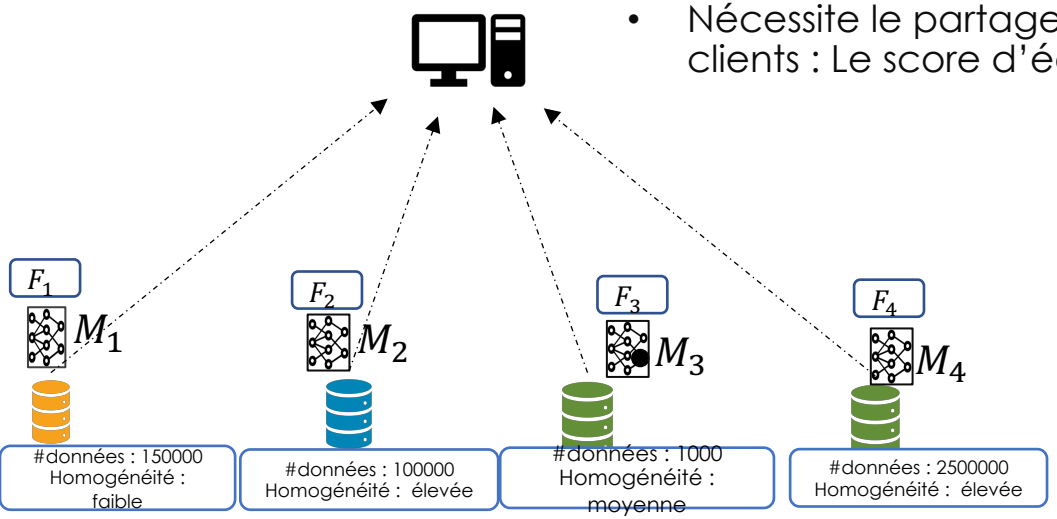
Conception d'une approche d'agrégation équitable et respectueuse de la vie privée pour l'apprentissage fédéré

Implémentation et évaluation des solutions proposées par rapport aux attaques identifiées.

Premiers résultats

$$M_{global} = \sum_{i=1}^{4} \omega_i \cdot M_i$$

$$\omega_i = f(F_i, \#données_i)$$



- Définition d'une fonction d'agrégation prenant en compte l'équité des modèles produits par les clients.
- Nécessite le partage d'une information supplémentaire par les clients : Le score d'équité de leur modèle (*)

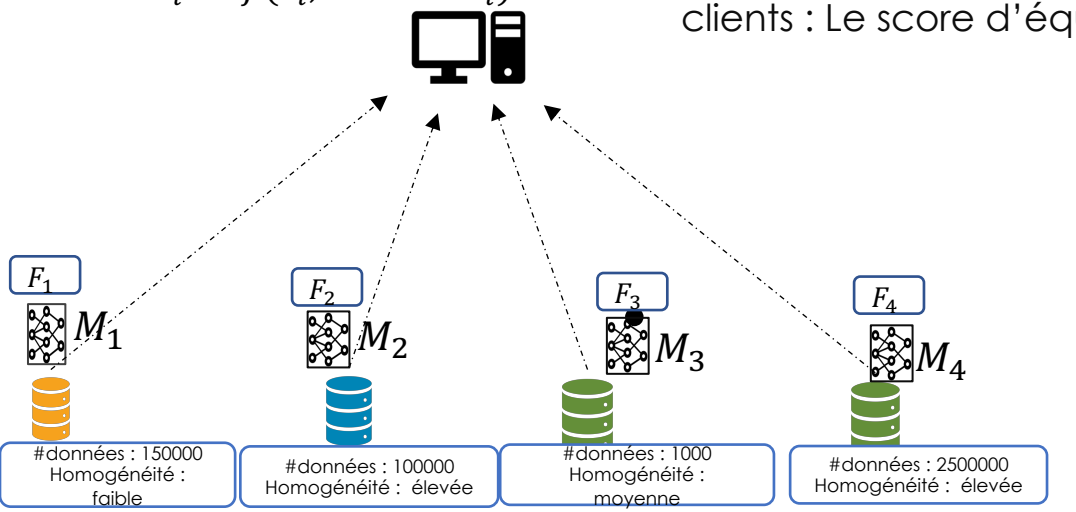
(*) Ezzeldin, Y.H., Yan, S., He, C., Ferrara, E. and Avestimehr, S., 2021. Fairfed: Enabling group fairness in federated learning. *arXiv preprint arXiv:2110.00857*.

Premiers résultats

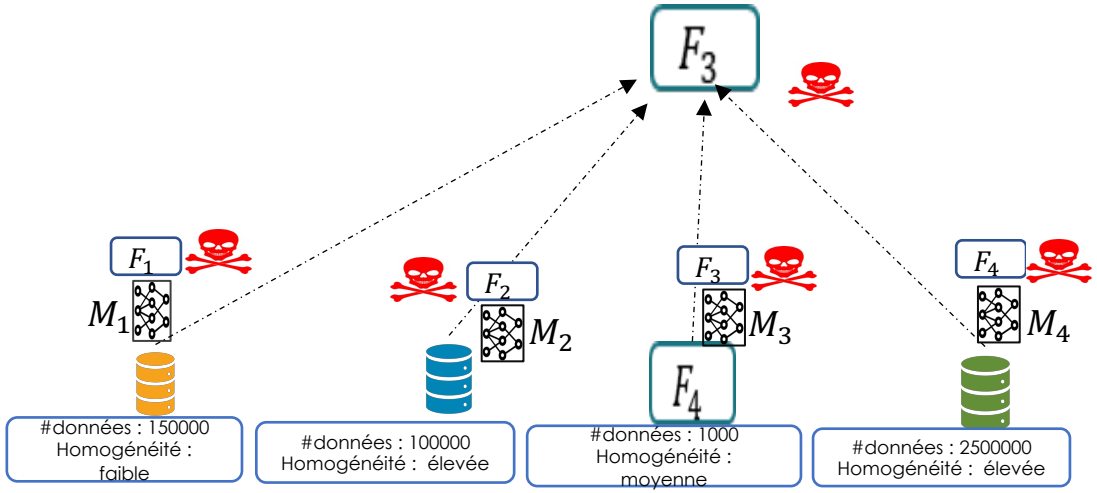
$$M_{global} = \sum_{i=1}^{4} \omega_i \cdot M_i$$

$$\omega_i = f(F_i, \#données_i)$$

- Définition d'une fonction d'agrégation prenant en compte l'équité des modèles produits par les clients.
- Nécessite le partage d'une information supplémentaire par les clients : Le score d'équité de leur modèle



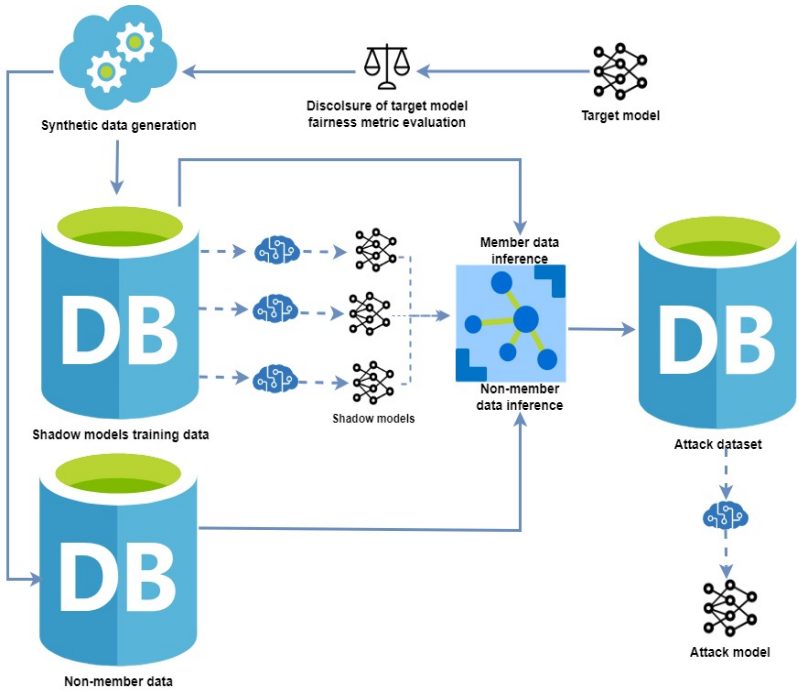
- La mesure d'équité peut être exploitée pour améliorer des attaques de reconstruction par un serveur d'agrégation honnête mais curieux (*)

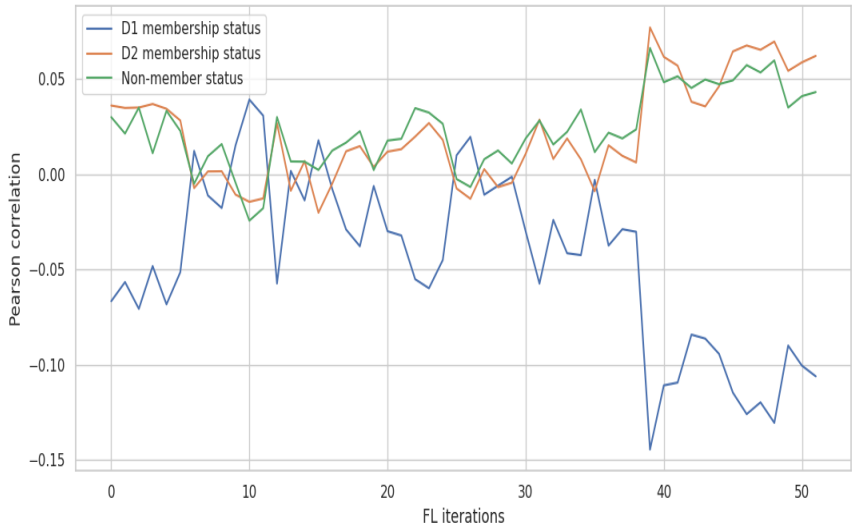


(*) Ferry, J., Aïvodji, U., Gambs, S., Huguet, M.J. and Siala, M., 2023, February. Exploiting Fairness to Enhance Sensitive Attributes Reconstruction. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)* (pp. 18-41).

Attaque par inférence améliorée

Améliorer l'équité d'un modèle aura pour conséquence l'augmentation des risques des attaques d'inférence (membership/attribute inference) sur la base de données associée.



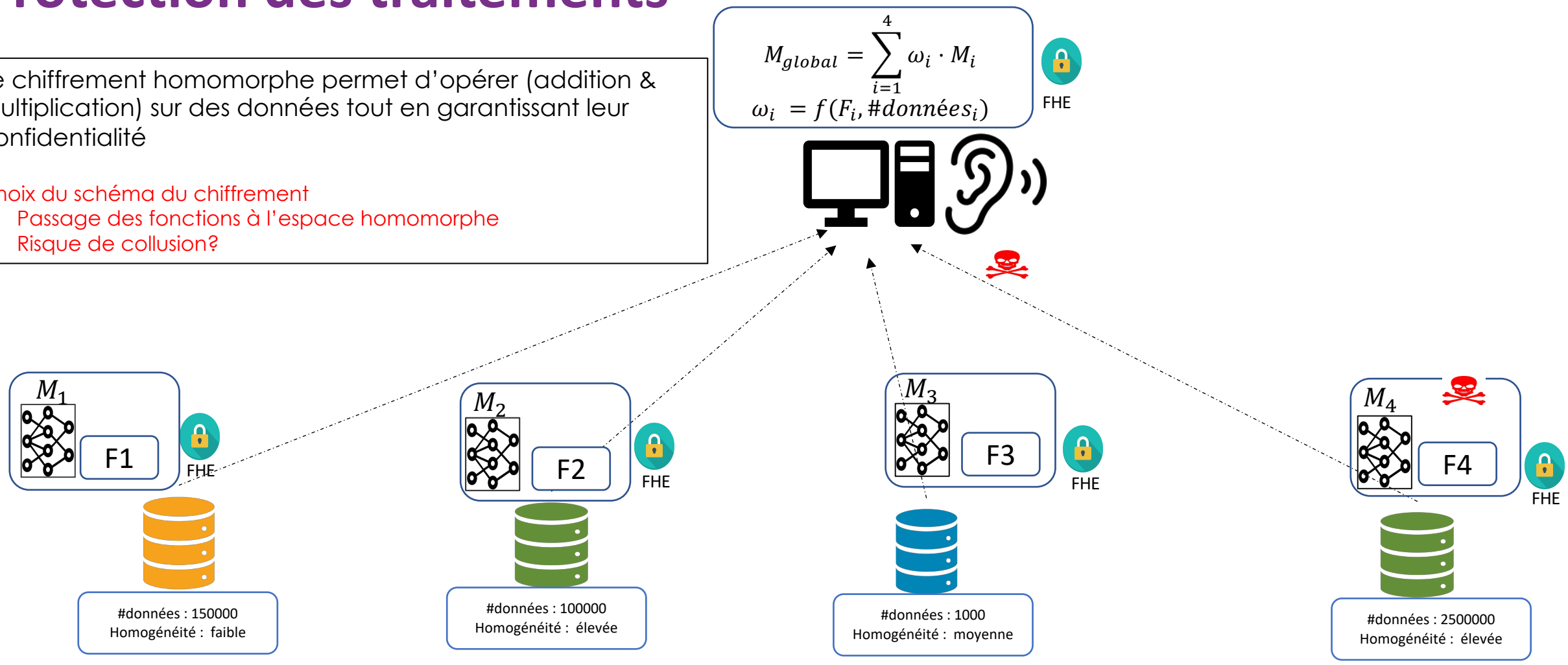


Distance entre les prédictions du modèle agrégé et les prédictions de deux modèles provenant de deux data-sets en fonction de leurs poids respectifs.

Protection des traitements

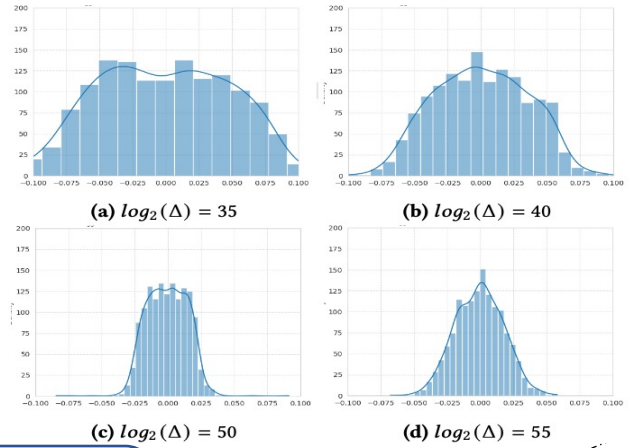
Le chiffrement homomorphe permet d'opérer (addition & multiplication) sur des données tout en garantissant leur confidentialité

Choix du schéma du chiffrement
⇒ Passage des fonctions à l'espace homomorphe
⇒ Risque de collusion?



Vers une protection contre les collisions

Modélisation du bruit engendré par CKKS – Garantie de la DP (*)



(a) $log_2(\Delta) = 35$ (b) $log_2(\Delta) = 40$

(c) $log_2(\Delta) = 50$ (d) $log_2(\Delta) = 55$

$$M_{global} = \sum_{i=1}^{4} \omega_i \cdot M_i$$

$$\omega_i = f(F_i, \#données_i)$$

FHE

M_1 F1 FHE

#données : 150000
Homogénéité : faible

M_2 F2 FHE

#données : 100000
Homogénéité : élevée

M_3 F3 FHE

#données : 1000
Homogénéité : moyenne

M_4 F4 FHE

#données : 2500000
Homogénéité : élevée

EQUIHid : La santé numérique au service de tous

- **Résultats et productions:**
  - Intégration d'un consortium –international- de réseautage **CybAlliance** axé sur la sécurité et la vie privée dans le domaine de la santé.
  - Collaborer avec l'équipe PRIVATICS de l'INRIA, et Stevens Institute aux Etats Unis.
  - Inviter deux chercheurs du NTNU, Norvège et Stevens Institute, USA.

- **Secteur de la santé:**
  - Réponse à l'engagement de la France en tant qu'acteur majeur de la santé mondiale.
  - Soutien aux ambitions du 13ème programme général de l'OMS pour une vie saine et le bien-être à tous les âges.
  - Contribution à la conception d'applications IA équitables et protectrices des données médicales sensibles.

- **Secteur social:** Amélioration de la qualité des solutions de santé pour les citoyens.
  - Offre de services de santé personnalisés et adaptés à leurs besoins spécifiques.
  - Contribution à la recherche en santé éthique.

UQÀM

Sébastien Gambs
Didem Demirag

TELECOM SudParis
IP PARIS

Maryline Laurent
Akram Adda Bendoukha

Merci

cea

Renaud Sirdey
Aymen Boudguiga

relyens
sham | sofaxis

Abdallah Arioua

Medtronic
Engineering the extraordinary

Antoine Groheux

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Claire Levallois Barth

GENOPOLE
VIVRE L'INNOVATION

Victoria Del Angel