

Symposium Agorantic

ApprenTissage fédéré poUr deS données héTérogènes et sensibles (TRUST)

03-12-2024

Présenté par : **Moussaab SBAI**

moussaab.sbai@alumni.univ-avignon.fr



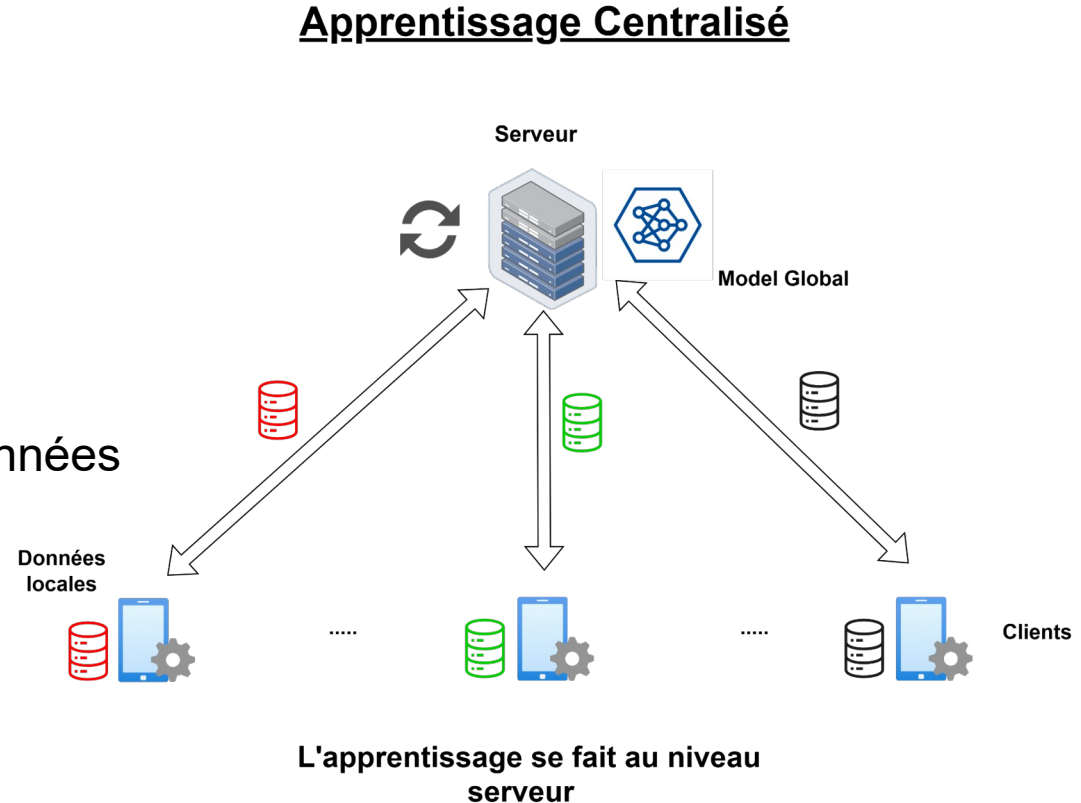
Apprentissage centralisé

- Approche classique des systèmes d'Intelligence Artificielle.

Gemini

ChatGPT

- Utilisation des quantités massives données générées par les utilisateurs 1,5 To

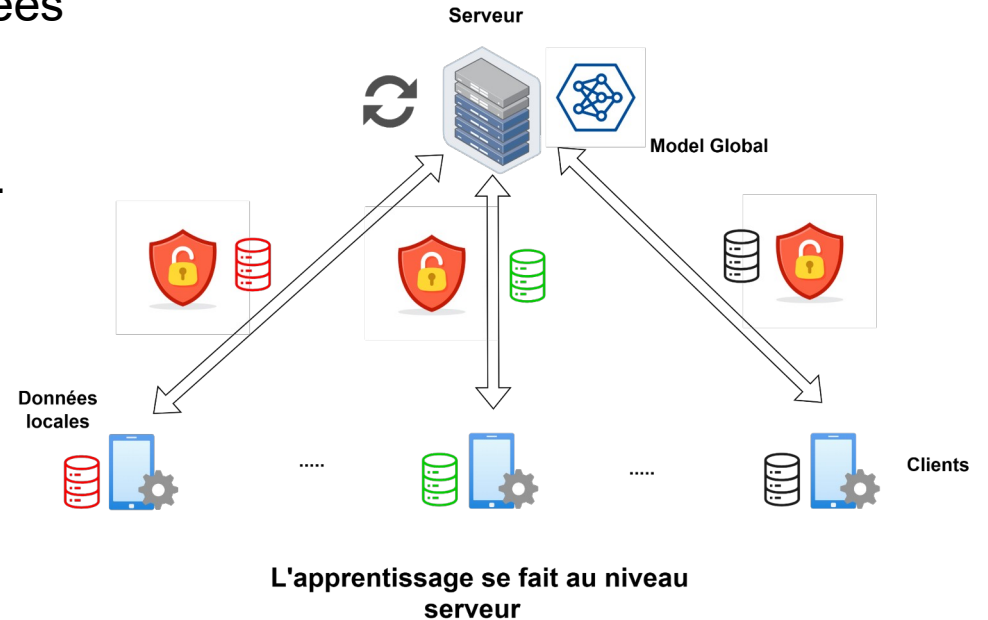


Apprentissage centralisé

Problématiques

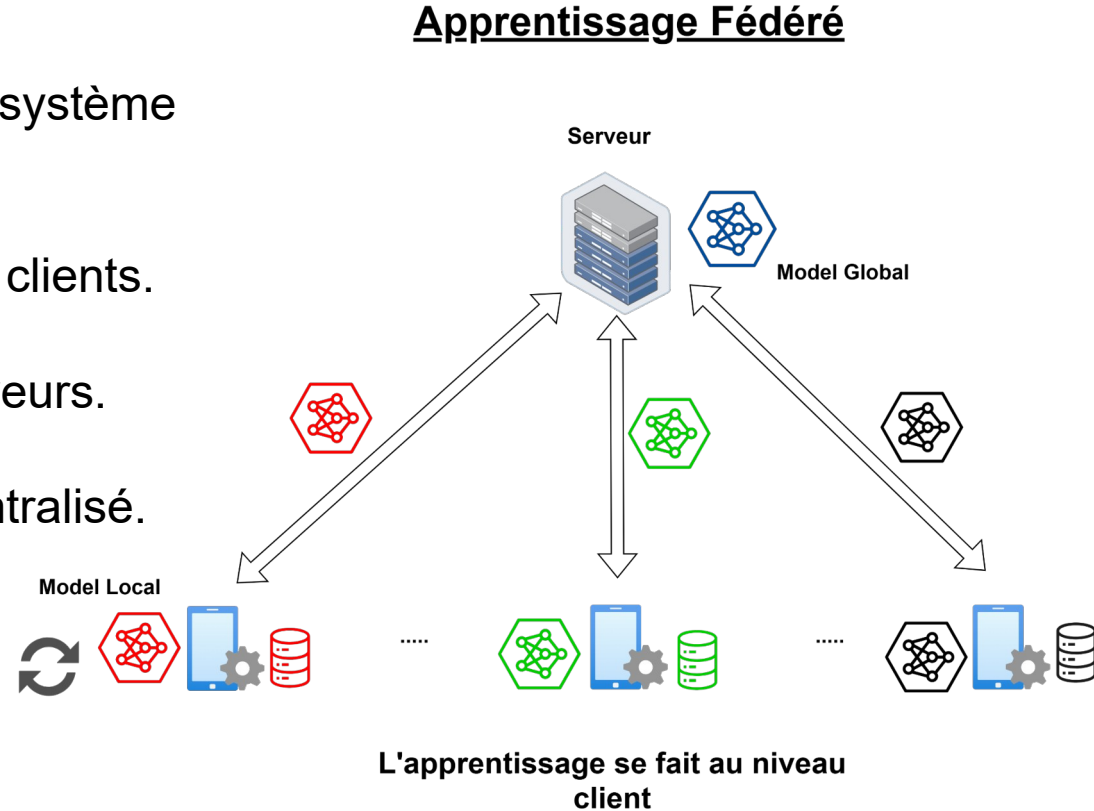
- Non-respect de la confidentialité des données de l'utilisateur.
- Besoin d'énormes ressources de stockage.
- Grande consommation d'énergie.
- Dépendance de serveur centralisé.

Apprentissage Centralisé



Apprentissage Fédéré

- Nouvelle approche d'apprentissage du système d'intelligence artificielle [1].
- L'apprentissage se fait au niveau des clients.
- Pas de transfert de données aux serveurs.
- Moins de dépendance au serveur centralisé.



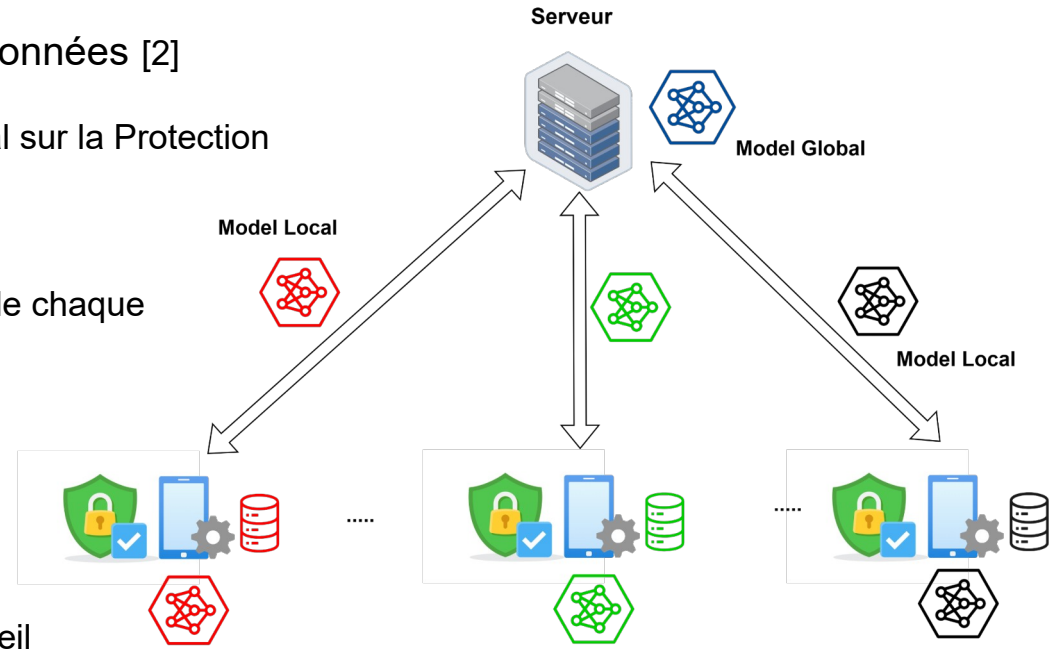
[1] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). *Communication-Efficient Learning of Deep Networks from Decentralized Data*. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, PMLR, Vol. 54, pp. 1273–1282.

Apprentissage Fédéré

Avantages

- Respecte de la confidentialité et de la sécurité des données [2]
 - Les données restent sur les appareils locaux
 - Conformité réglementaire : RGPD (Règlement Général sur la Protection des Données)
- Personnalisation de l'apprentissage
 - Les modèles résultant sont plus adaptés aux besoins de chaque utilisateur
- Réduction des coûts liés à la centralisation des données [3]
 - Pas de transfert de données massives dans le réseau
 - Pas de stockage centralisé des données massives
- Adapté à des environnements distribués
 - Moins sensible à une panne d'un nœud ou d'un appareil

Apprentissage Fédéré



[2] Li, Qinbin, et al. "A survey on federated learning systems: Vision, hype and reality for data privacy and protection." *IEEE Transactions on Knowledge and Data Engineering* 35.4 (2021): 3347-3366.

[3] Yang, Zhaohui, et al. "Energy efficient federated learning over wireless communication networks." *IEEE Transactions on Wireless Communications* 20.3 (2020): 1935-1949.

Apprentissage Fédéré

Domaines d'applications de l'Apprentissage Fédéré



Santé



Finance

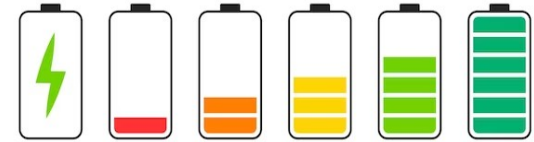
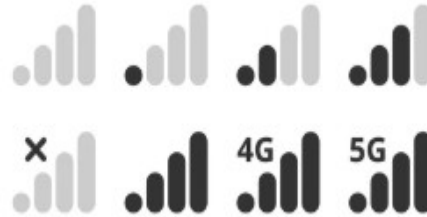
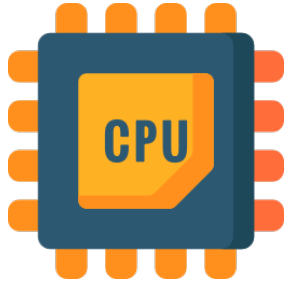


Gouvernement

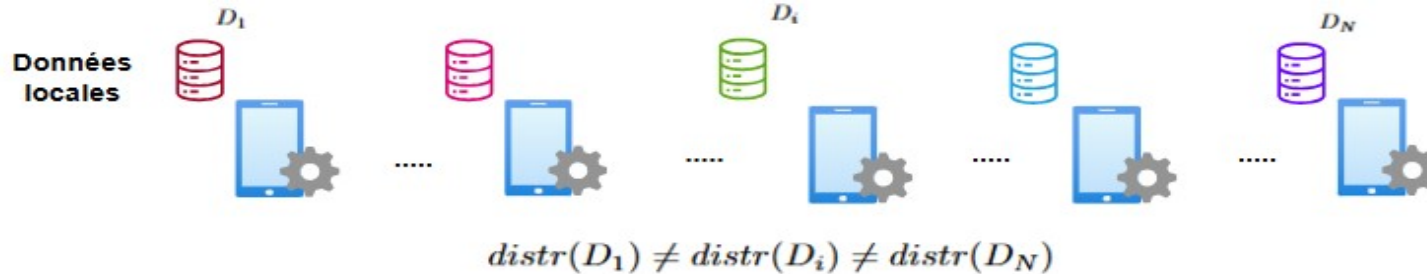
Motivations

Caractéristiques de l'Apprentissage Fédéré

Hétérogénéité système : Diversité des appareils participants



Hétérogénéité statistique :

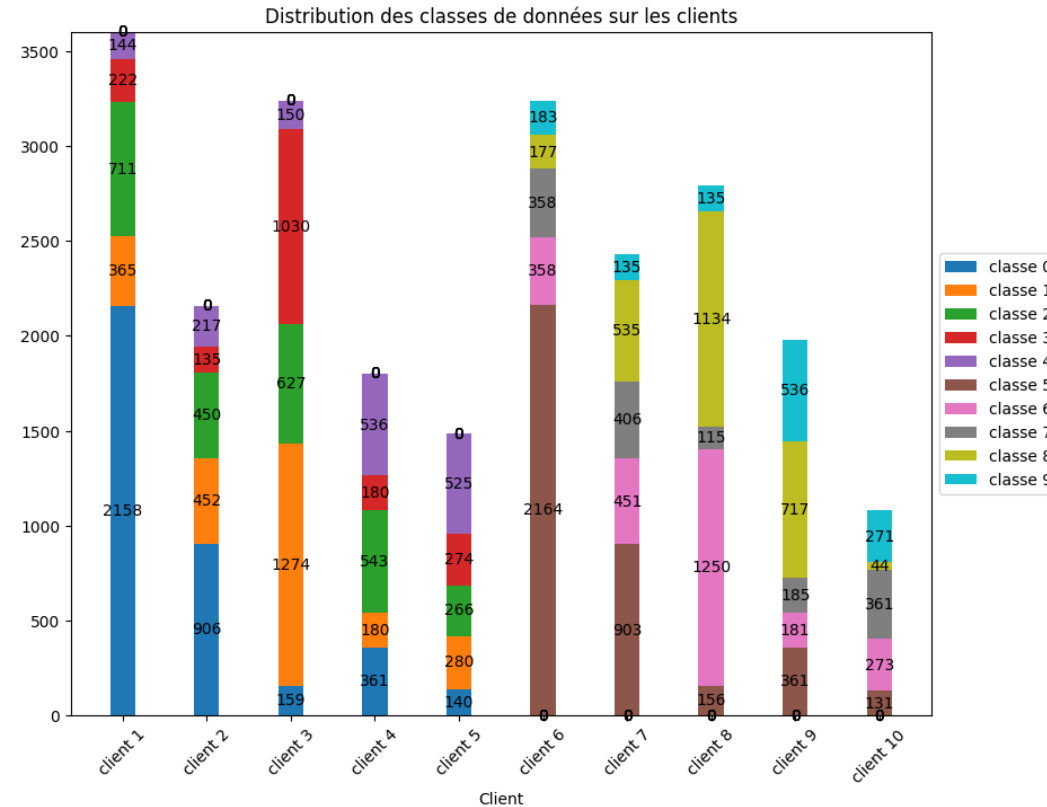


[4] Ye, Mang, et al. "Heterogeneous federated learning: State-of-the-art and research challenges." *ACM Computing Surveys* 56.3 (2023): 1-44.

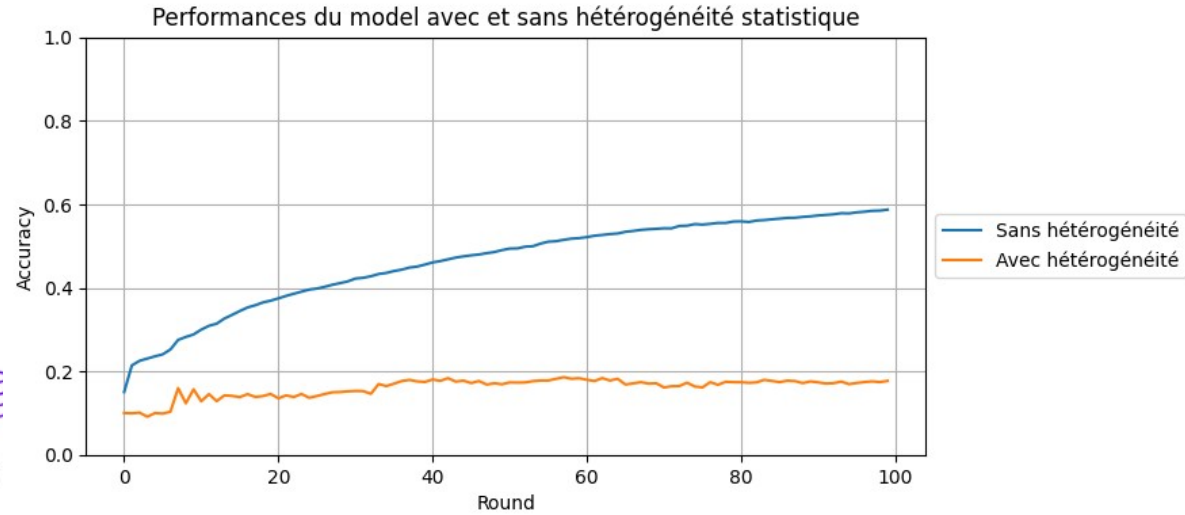
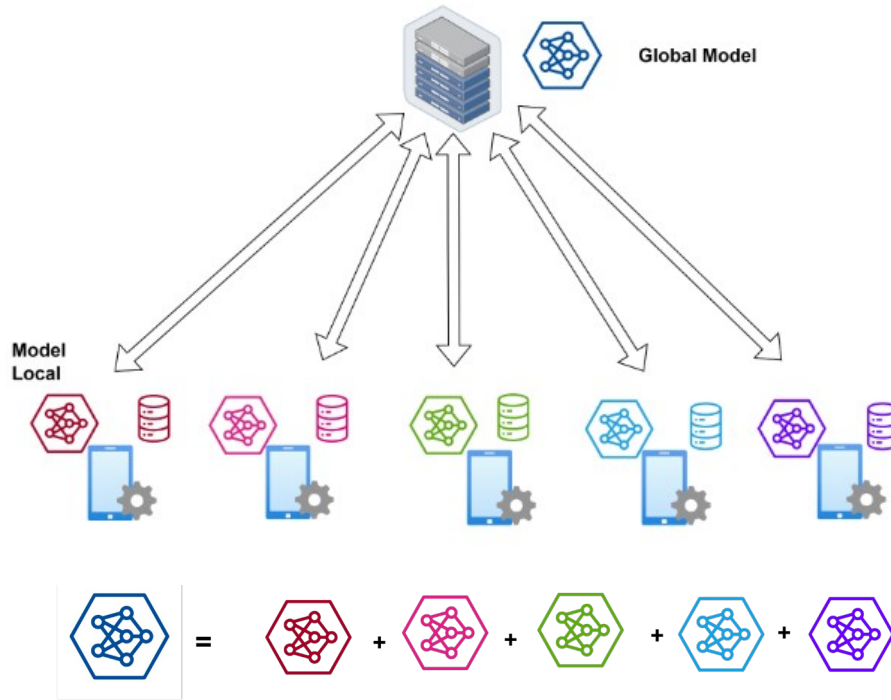
L'impact de l'hétérogénéité statistique sur l'apprentissage fédéré.

Illustration de l'hétérogénéité statistique

- Les distributions des classes des données locales des clients sont différentes



L'impact de l'hétérogénéité statistique sur l'apprentissage fédéré.



Affecte les performances du model Global

[4] Ye, Mang, et al. "Heterogeneous federated learning: State-of-the-art and research challenges." *ACM Computing Surveys* 56.3 (2023): 1-44.

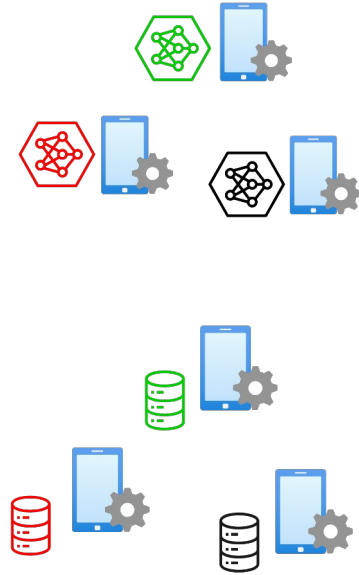
Problématiques

- **Quantifier la contribution statistique de chaque client tout en garantissant la confidentialité de leurs données.**
- **Concevoir des schémas intelligents de sélection des clients.**
- **Développer des stratégies efficaces d'entraînement des clients.**

Quantification de la contribution statistique de chaque client

Contribution ou utilité statistique

- Approches basées sur les modèles des clients :
 - La divergence des modèles locaux par rapport au modèle global [5].
 - Le pourcentage de similarité des signes entre les modèles locaux et le modèle global.
- Approches basées sur les données des clients :
 - Plus d'importance aux clients avec plus de données.
 - Mesures statistiques des données locales des clients [6].
 - Peut porter atteinte à la confidentialité des données.



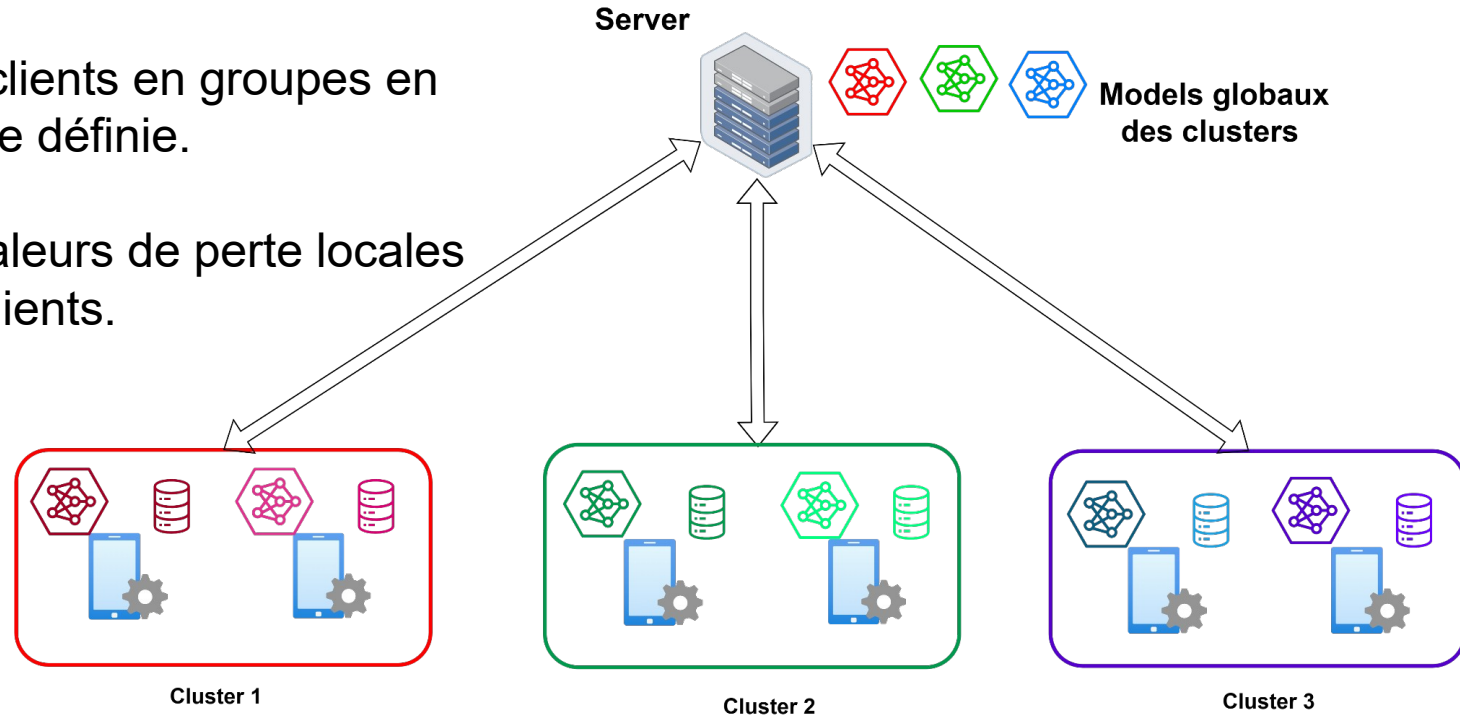
[5]. Xiao Zeng, Ming Yan, and Mi Zhang. Mercury: Efficient on-device distributed dnn training via stochastic importance sampling. In Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, pages 29–41, 2021.

[6]. Anass Houdou, Hamza Alami, Khalid Fardousse, Ismail Berrada, et al. Nifl: A statistical measures-based method for client selection in federated learning. IEEE Access, 10:124766–124776, 2022.

Schémas intelligents de sélection des clients

Clustering des clients

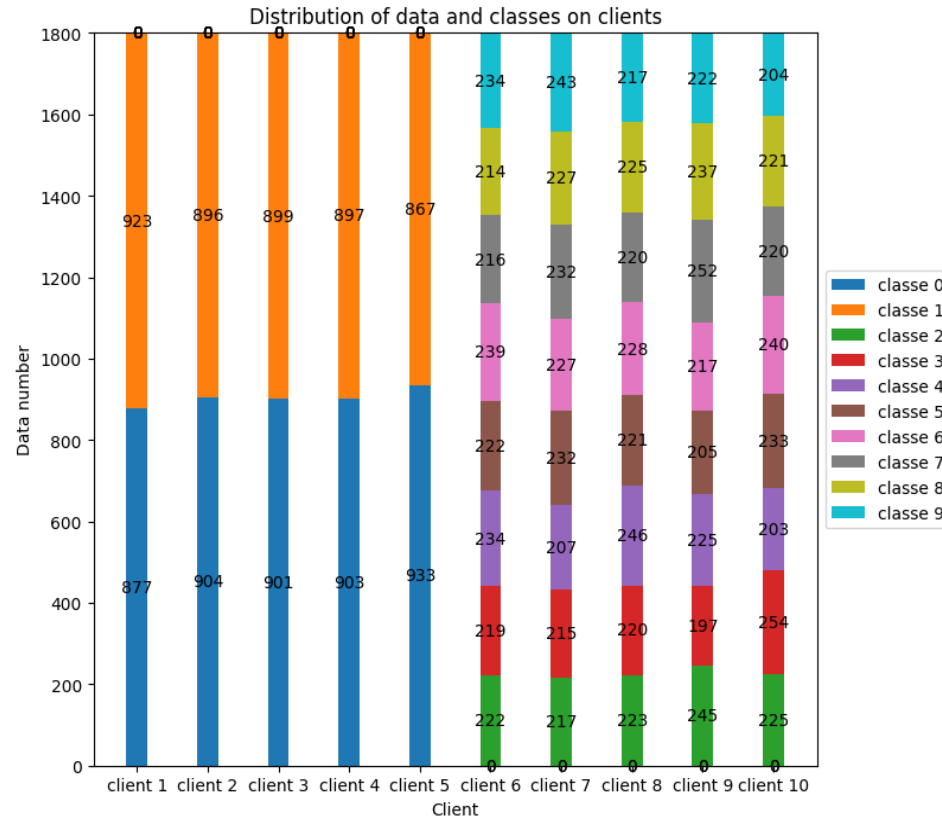
- Regroupement des clients en groupes en fonction une métrique définie.
- IFCA [7] utilise les valeurs de perte locales pour regrouper les clients.



[7]. Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. Advances in Neural Information Processing Systems, 33:19586– 19597, 2020.

Schémas intelligents de sélection des clients

Scénario : hétérogénéité des données entre groupes.

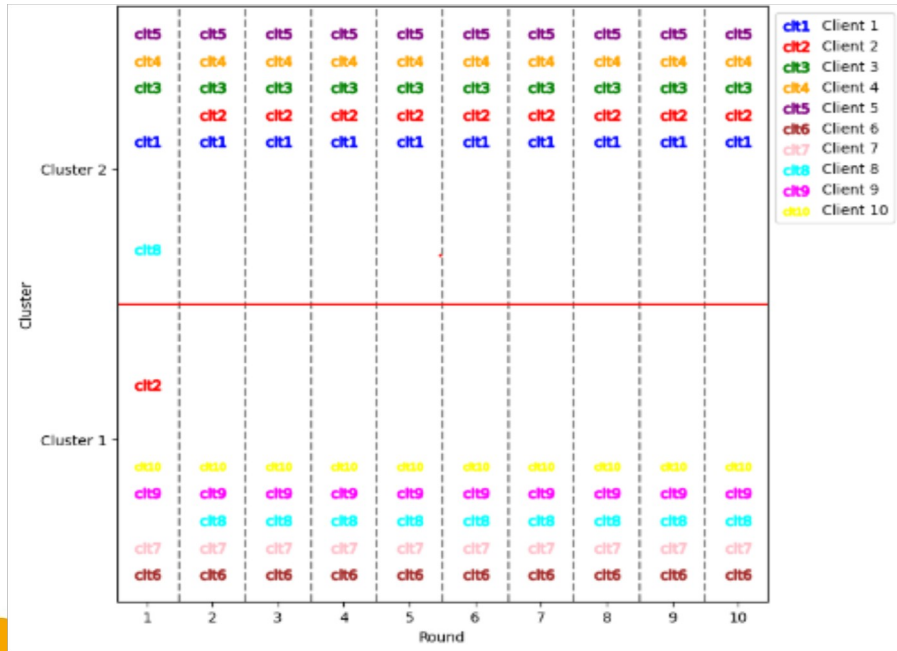


Schémas intelligents de sélection des clients

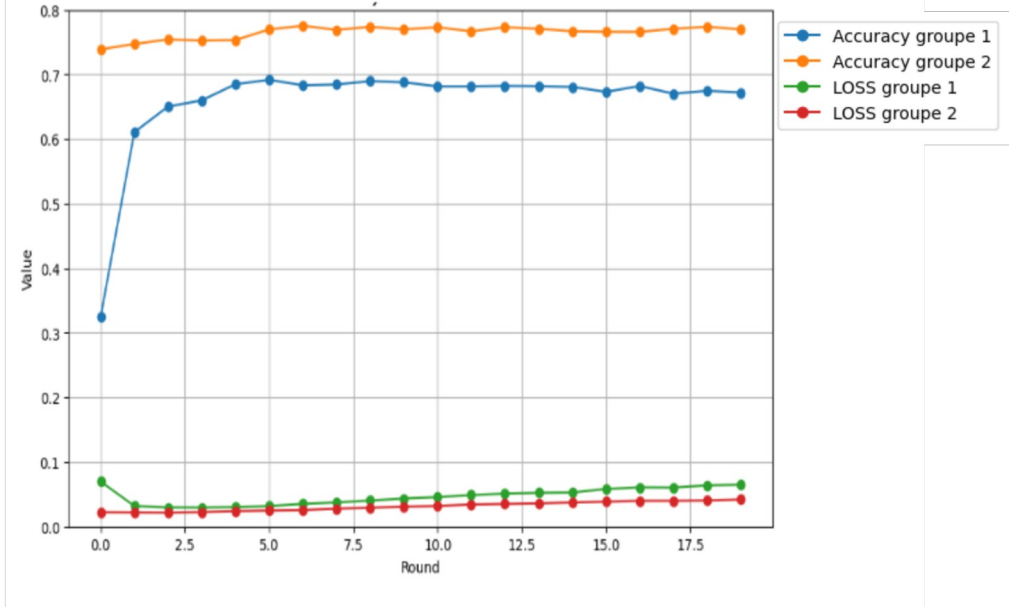
Clustering des clients

Résultats des expériences

Résultats de la répartition des clients sur les clusters



Performance des modèles résultant du clustering

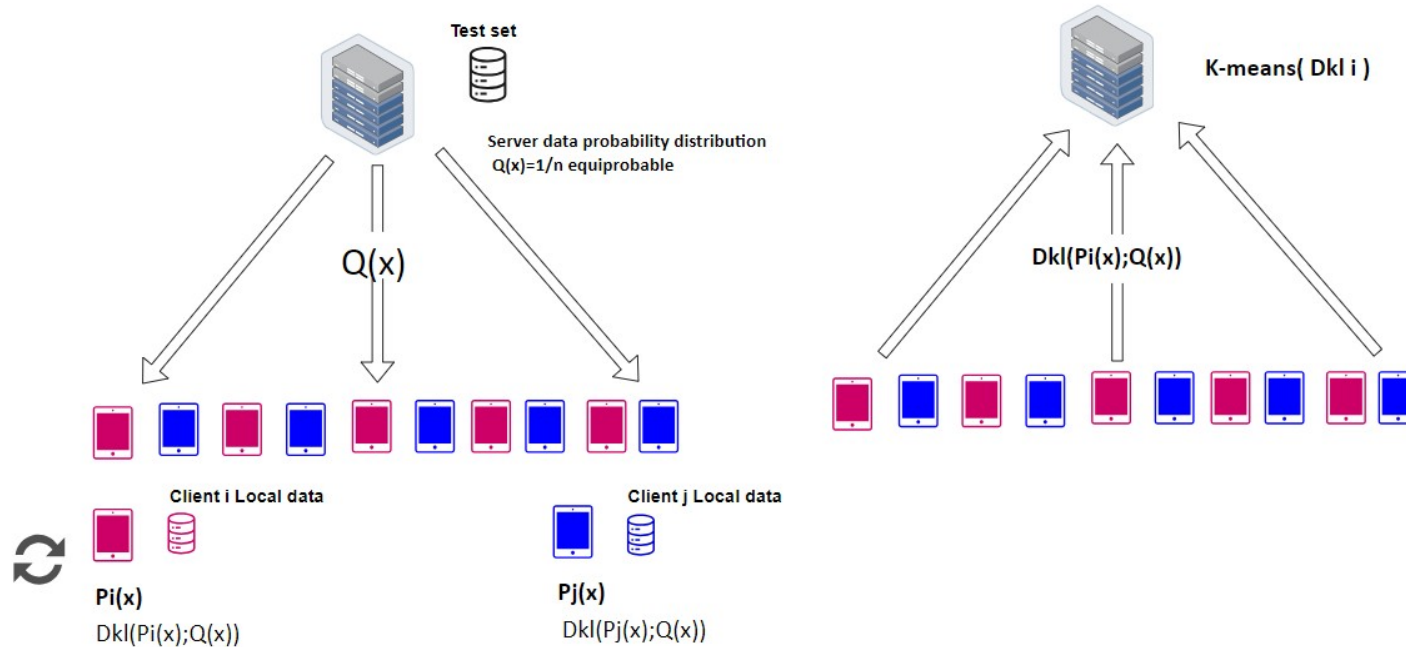


Schémas intelligents de sélection des clients

Clustering des clients

Méthode proposée

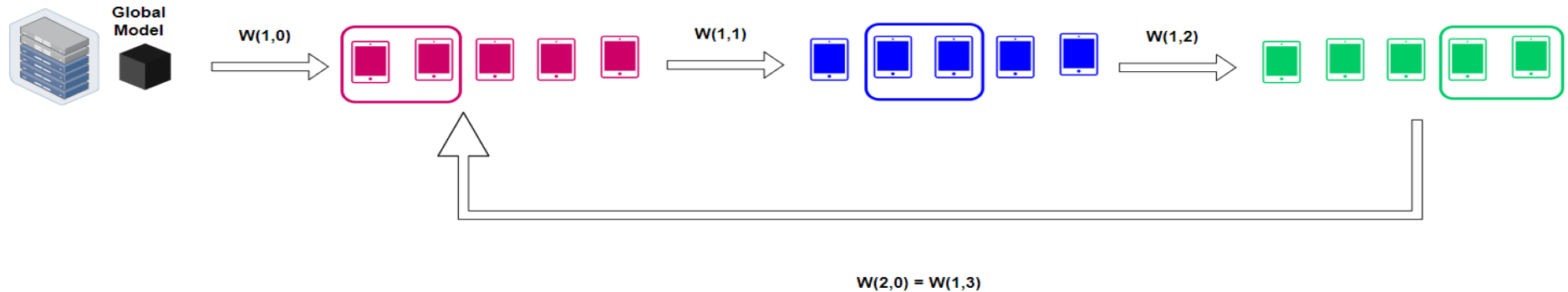
Former des groupes de clients à l'aide de la divergence de **Kullback-Leibler (KL)**, des distributions de probabilité des classes des données des clients et des algorithmes de clustering.



Stratégies efficaces d'entraînement des clients

Participation cyclique des clients

- Ce travail [8] introduit une nouvelle méthode de formation du modèle global.
- Les clients sont regroupés en groupes prédéfinis



[8]. Yae Jee Cho, Pranay Sharma, Gauri Joshi, Zheng Xu, Satyen Kale, and Tong Zhang. On the convergence of federated averaging with cyclic client participation. In International Conference on Machine Learning, pages 5677–5721. PMLR, 2023.

Stratégies efficaces d'entraînement des clients

Participation cyclique des clients

Expériences

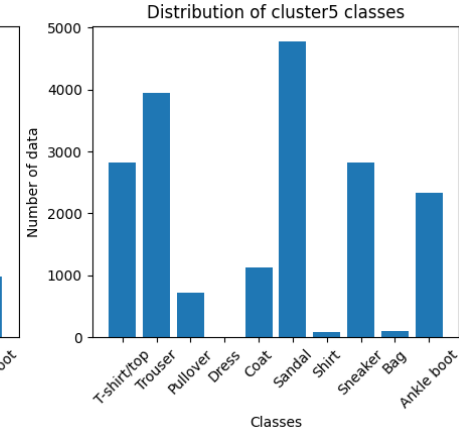
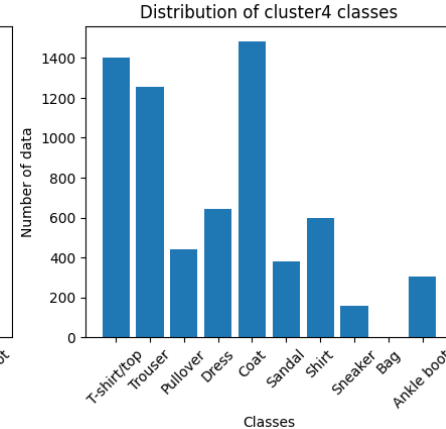
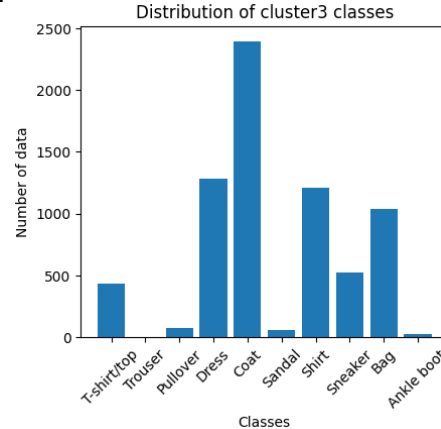
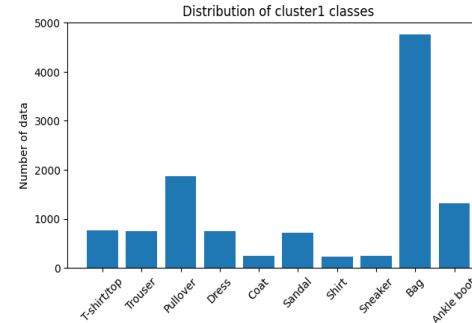
- **L'influence du taux de participation des clients.**
- **L'impact de la qualité du groupe de clients :**
 - Variation de l'hétérogénéité des données intra-groupe
 - Variation de l'hétérogénéité des données entre les groupes

Stratégies efficaces d'entraînement des clients

Participation cyclique des clients

Configuration des expériences

- 100 nombre total de clients.
- 5 distributions de classes des données différentes.
- 5 groupes contenant 20 clients avec une distribution de données similaire

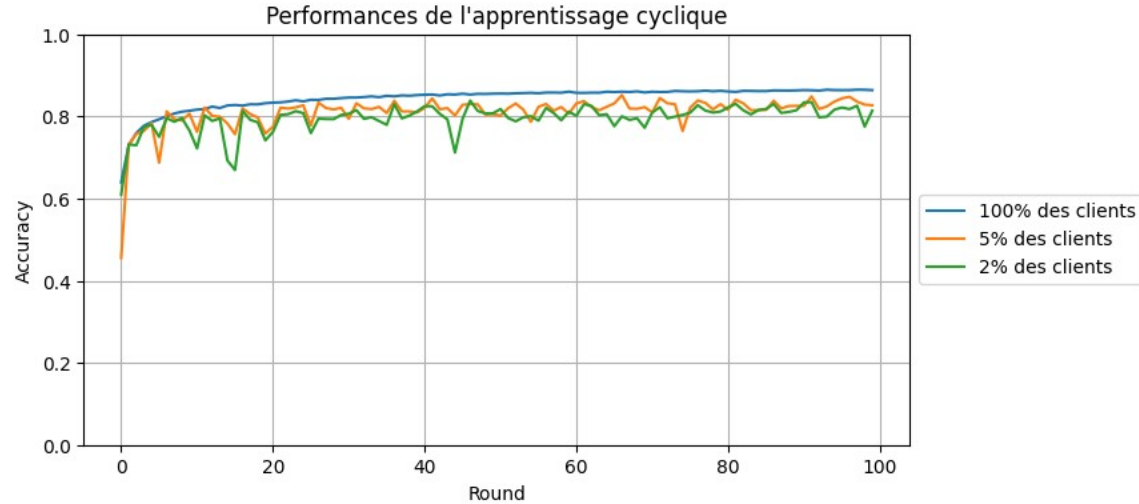


Stratégies efficaces d'entraînement des clients

Participation cyclique des clients

Résultats des expériences sur L'influence du taux de participation des clients

- 5% et 2% des clients participent l'entraînement du models



Conclusions des expériences

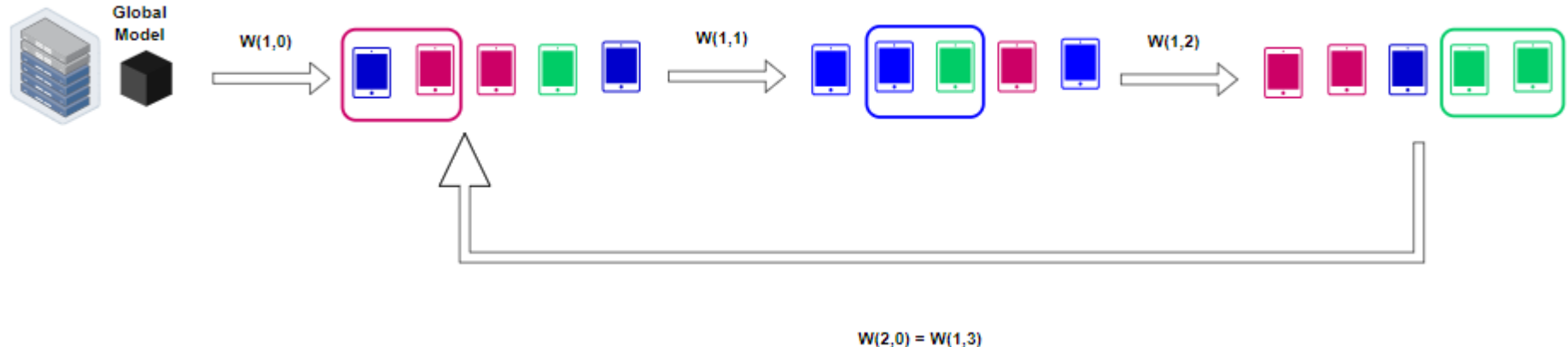
- Le taux de participation des clients n'ont pas un impact remarquable sur la performance du modèle.

Stratégies efficaces d'entraînement des clients

Participation cyclique des clients

L'impact de la qualité du groupe de clients

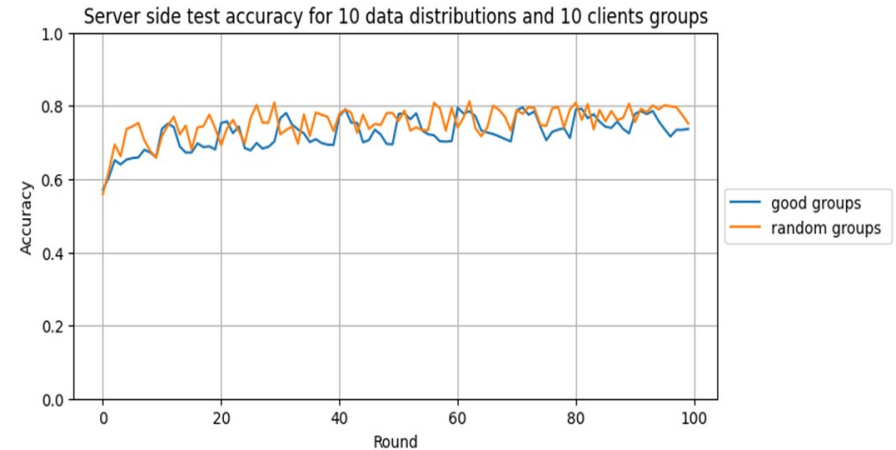
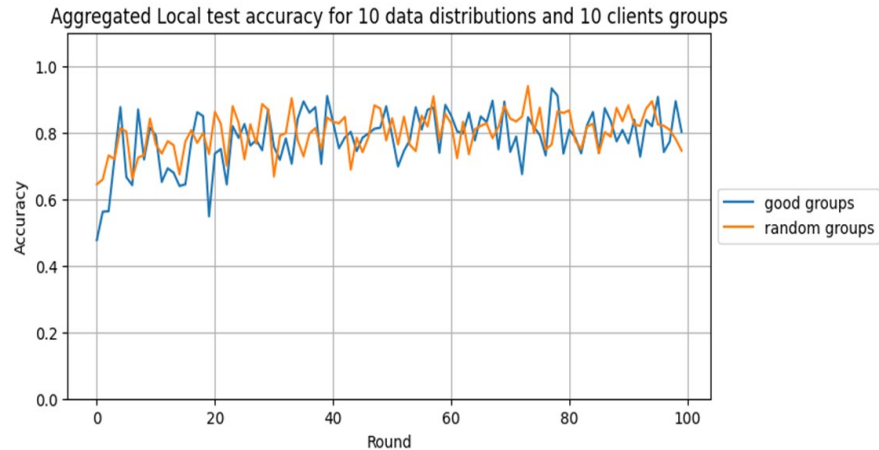
- Augmenter l'hétérogénéité des données au sein des groupes
- Au lieu de former des groupes de clients en fonction de la répartition de leurs données, les groupes sont formés de manière aléatoire.



Stratégies efficaces d'entraînement des clients

Participation cyclique des clients

Résultats des expériences sur L'impact de la qualité du groupe de clients



Conclusions des expériences

- La structure et la qualité des groupes de clients a pas d'impact sur la performance du modèle.

Stratégies efficaces d'entraînement des clients

Participation cyclique des clients

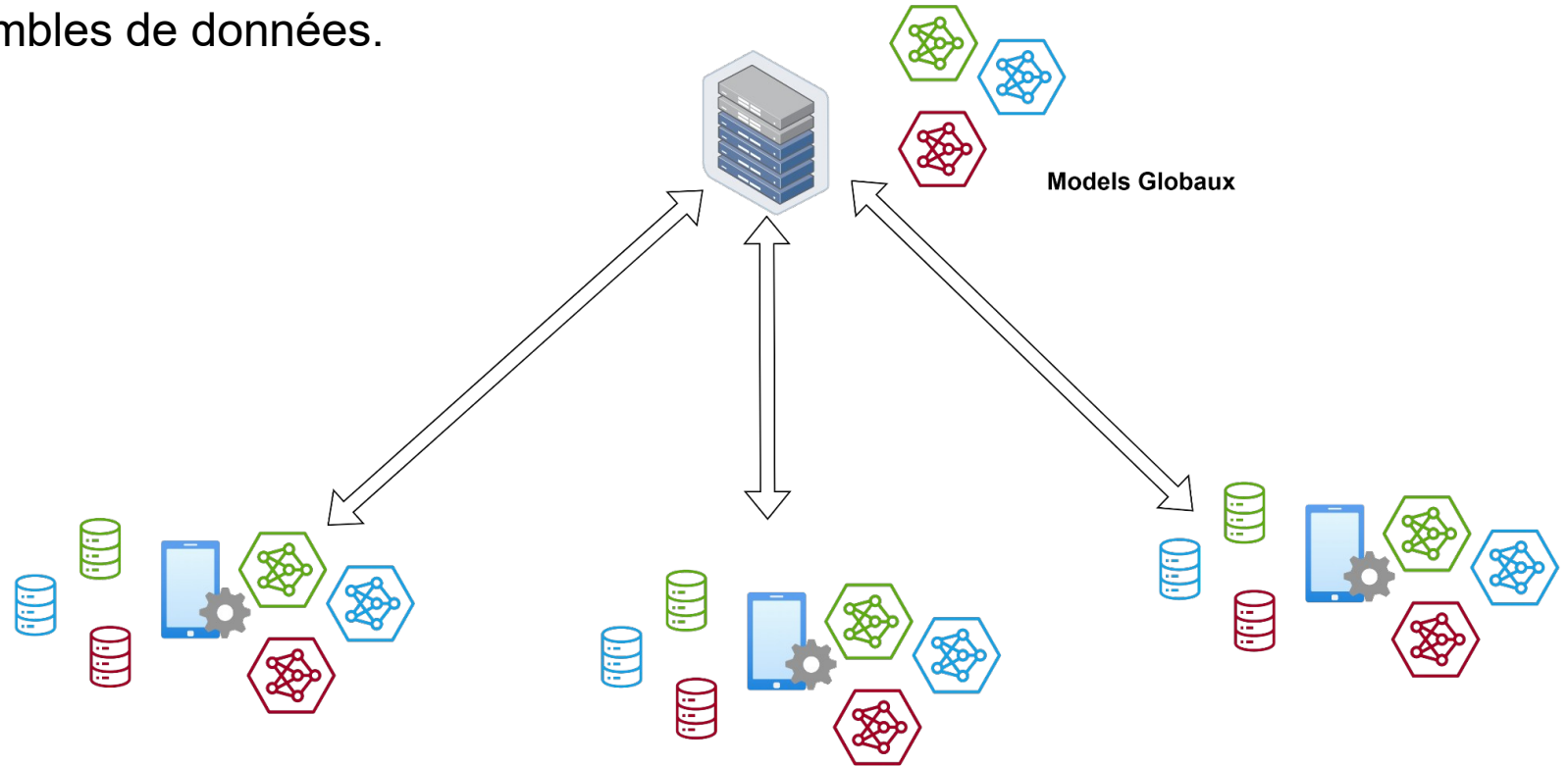
Conclusions des expériences

- La structure et la qualité des groupes de clients et les taux de participation des clients n'ont pas d'impact sur la performance du modèle.
- Nous étudions l'analyse théorique sous-jacente des modèles d'apprentissage fédéré pour proposer de nouvelles configurations pour les tests

Travaux en cours

Apprentissage fédéré à modèles multiples (MMFL)

Les clients participent à la formation de plusieurs modèles avec différentes architectures, tâches et ensembles de données.



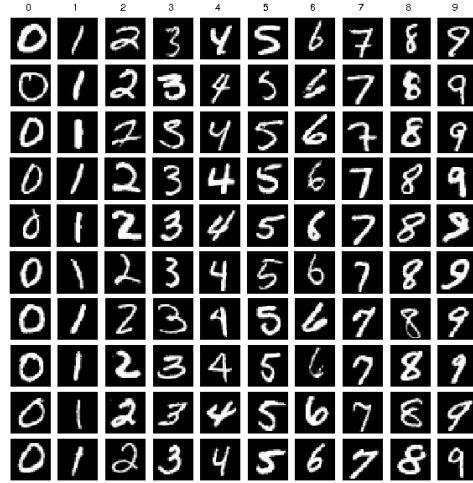
Conclusion

- L'apprentissage fédéré est une approche innovante pour maintenir la confidentialité des données.
- L'utilisation des données depuis plusieurs appareils expose l'apprentissage fédéré aux défis liés à des distributions très variables des données qui affecte ses performances.
- Étudier l'impact de l'hétérogénéité statistique sur le modèle global de l'apprentissage fédéré et proposer des solutions efficaces pour améliorer ses performances

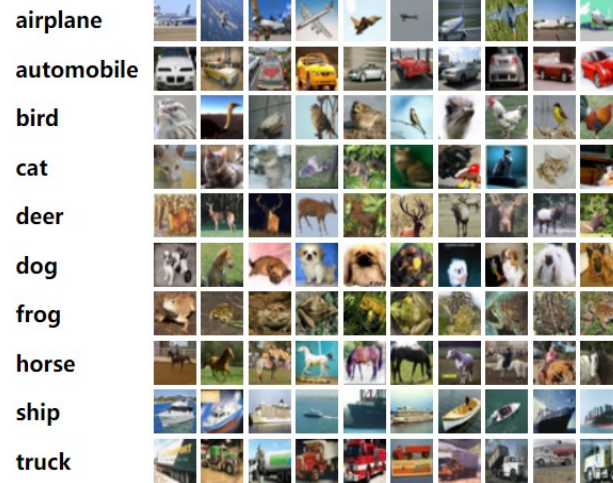
Merci.

Schémas intelligents de sélection des clients

Bases de données utilisées



MNIST



CIFAR10



FashionNIST

Apprentissage fédéré à modèles multiples (MMFL)

Challenges :

- Créer des groupes de clients unifiés capables d'améliorer efficacement les performances globales de tous les modèles en gérant l'hétérogénéité des données dans tous les ensembles de données locaux pour chaque modèle.
- Former des groupes de clients qui conviennent à tous les modèles.
- Développer une méthode d'allocation des clients qui optimise la consommation des ressources tout en maximisant les performances de chaque modèle.
- MMFL Nécessite des stratégies pour définir les priorités entre les modèles.
- Proposer un schéma de formation qui maintient l'équité entre les modèles, en équilibrant les performances et l'allocation des ressources sur l'ensemble du système.

[9]. Marie Siew, Haoran Zhang, Jong-Ik Park, Yuezhou Liu, Yichen Ruan, Lili Su, Stratis Ioannidis, Edmund Yeh, and Carlee Joe-Wong. Fair concurrent training of multiple models in federated learning. arXiv preprint arXiv:2404.13841, 2024.