cetic
Your Connection to ICT Research

Exploitez vos données d'entreprise sans les exposer

- Apprentissage fédéré au service d'une IA confidentielle -

Cyberwal by digital wallonia

WALHUB
BOOST YOUR DIGITAL TRANSFORMATION

Cyberweek @ A6K
28-11-2025

Xavier Lessage, Ir, PhD
Expert researcher (IA, Cyber) @ CETIC

www.enmieux.be

FEDER
UNION EUROPEENNE
Wallonie
LE FONDS EUROPÉEN DE DÉVELOPPEMENT RÉGIONAL ET LA WALLONIE INVESTISSENT DANS VOTRE AVENIR

official partner | digital wallonia .be

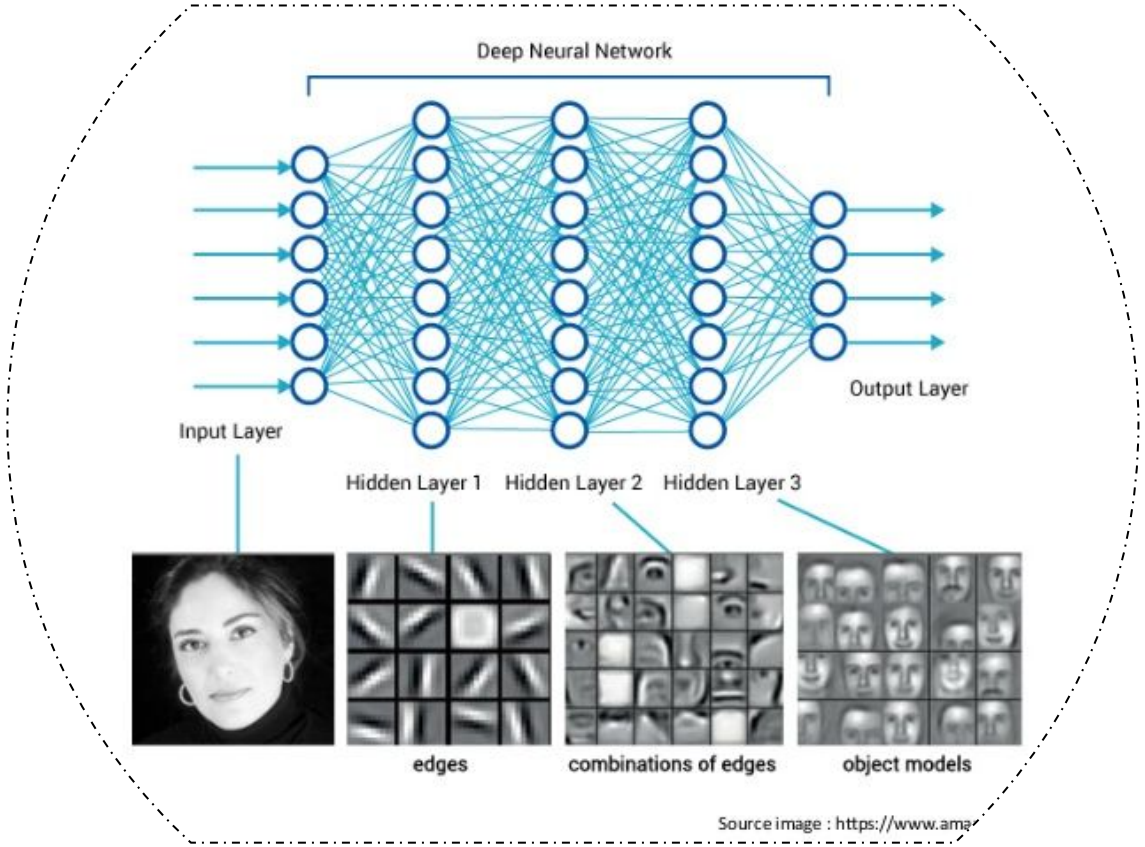
BOSA

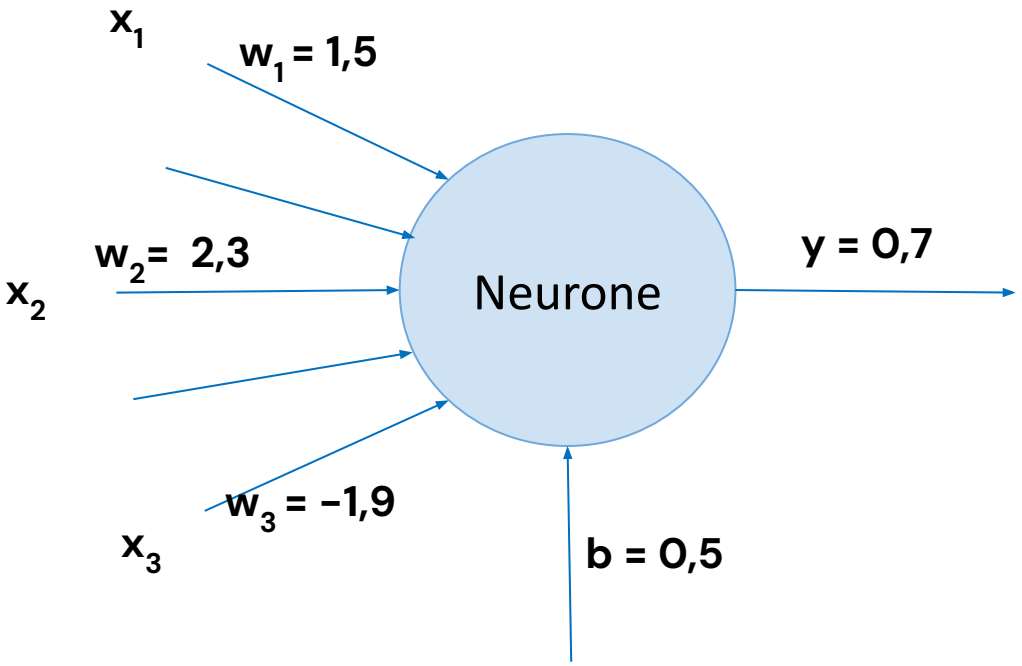
Centre d'Excellence en **Technologies** de l'**Information** et de la **Communication**
www.cetic.be

Agenda

- Apprentissage traditionnel
- Apprentissage fédéré
- Risques de sécurité liés à l'apprentissage fédéré
- Apprentissage fédéré sécurisé
  - Confidentialité différentielle appliquée au modèle
  - Chiffrement homomorphe appliqué au modèle
- Démo / Apprentissage fédéré sécurisé
  - Confidentialité différentielle
  - Chiffrement homomorphe
- Questions

Apprentissage traditionnel | Modèle



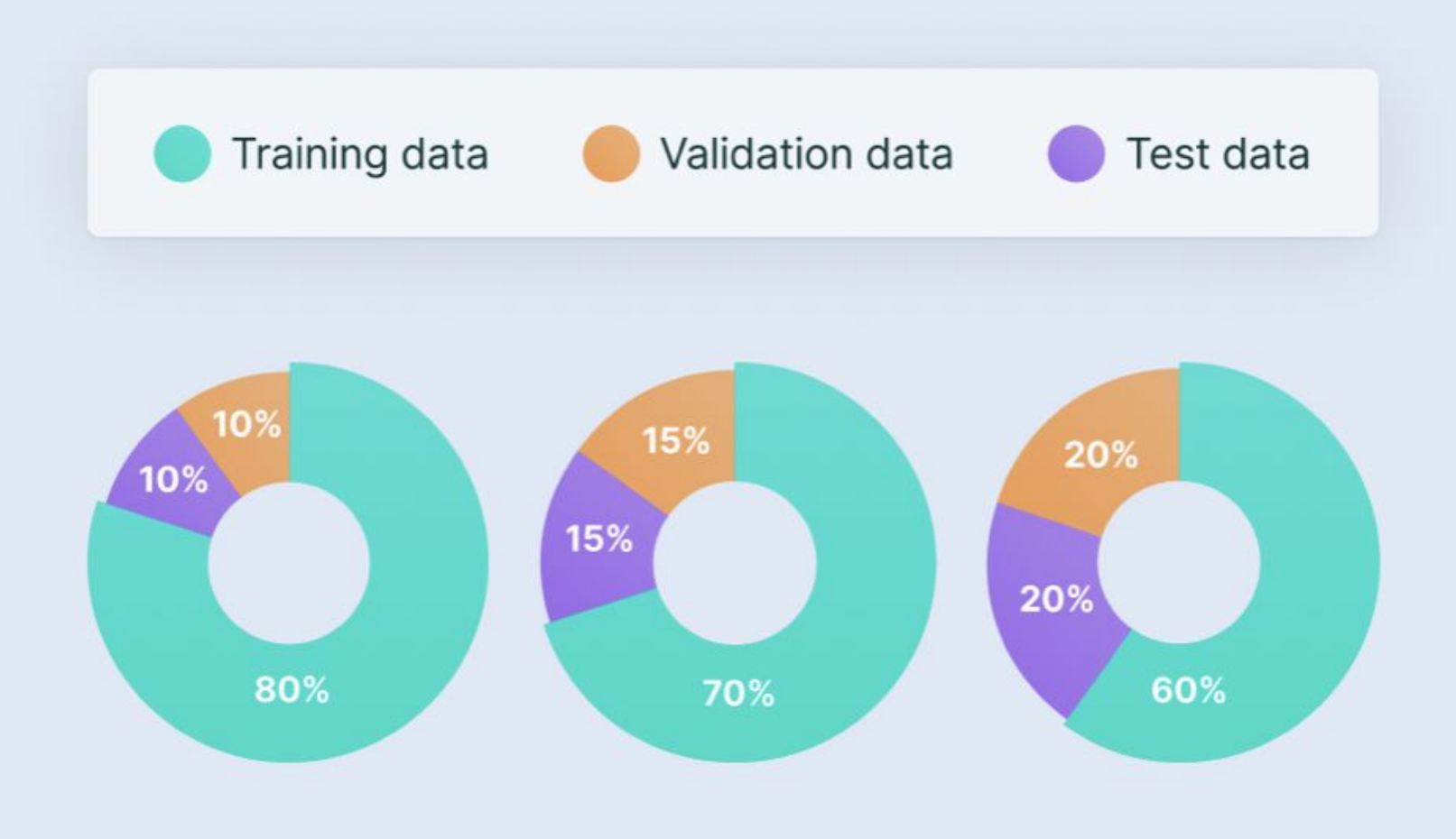


Deep learning → Basé du fonctionnement de notre cortex visuel

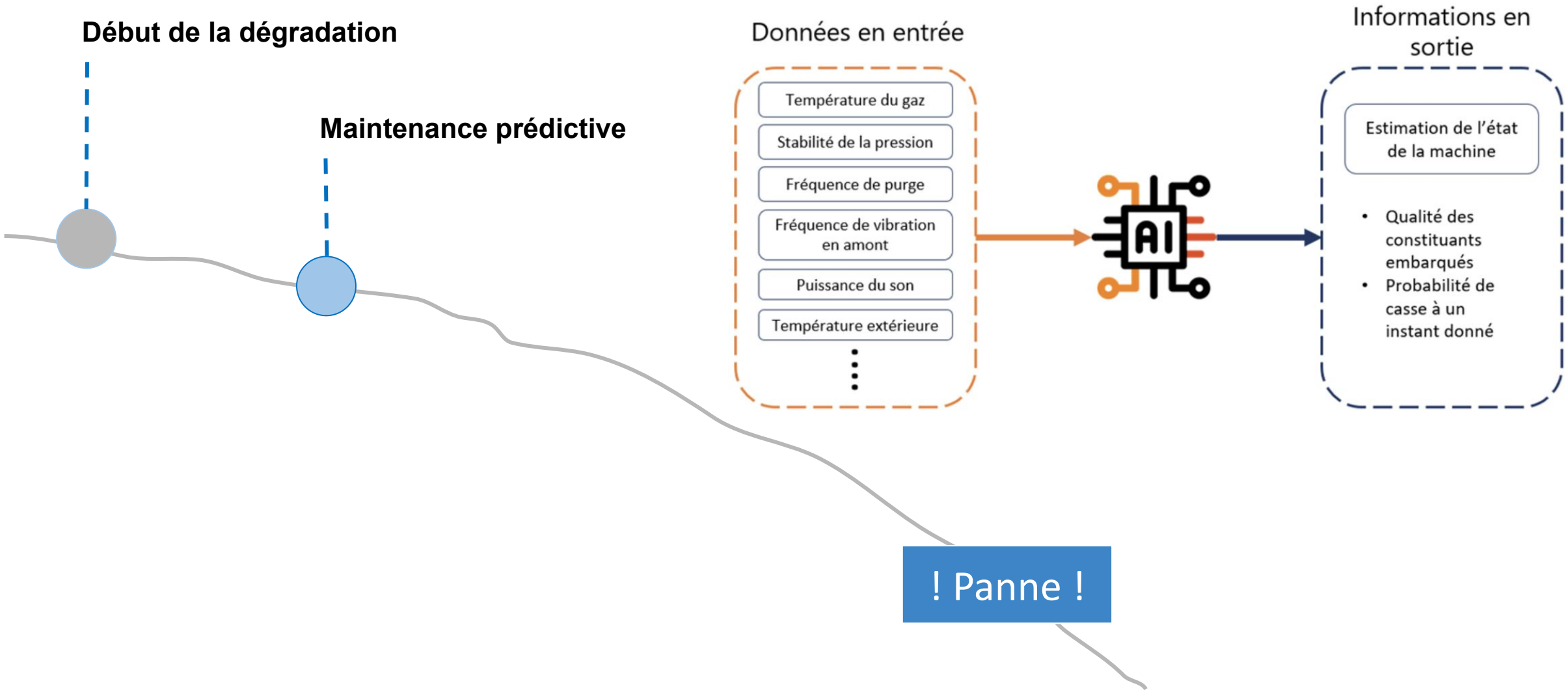
- Neurones sensibles à certaines zones / caractéristiques élémentaires
- Neurones sensibles à la combinaison de ces zones

$$Y_i = f\left(\sum_{j=1}^{n} W_{ij} X_j + B_i\right)$$

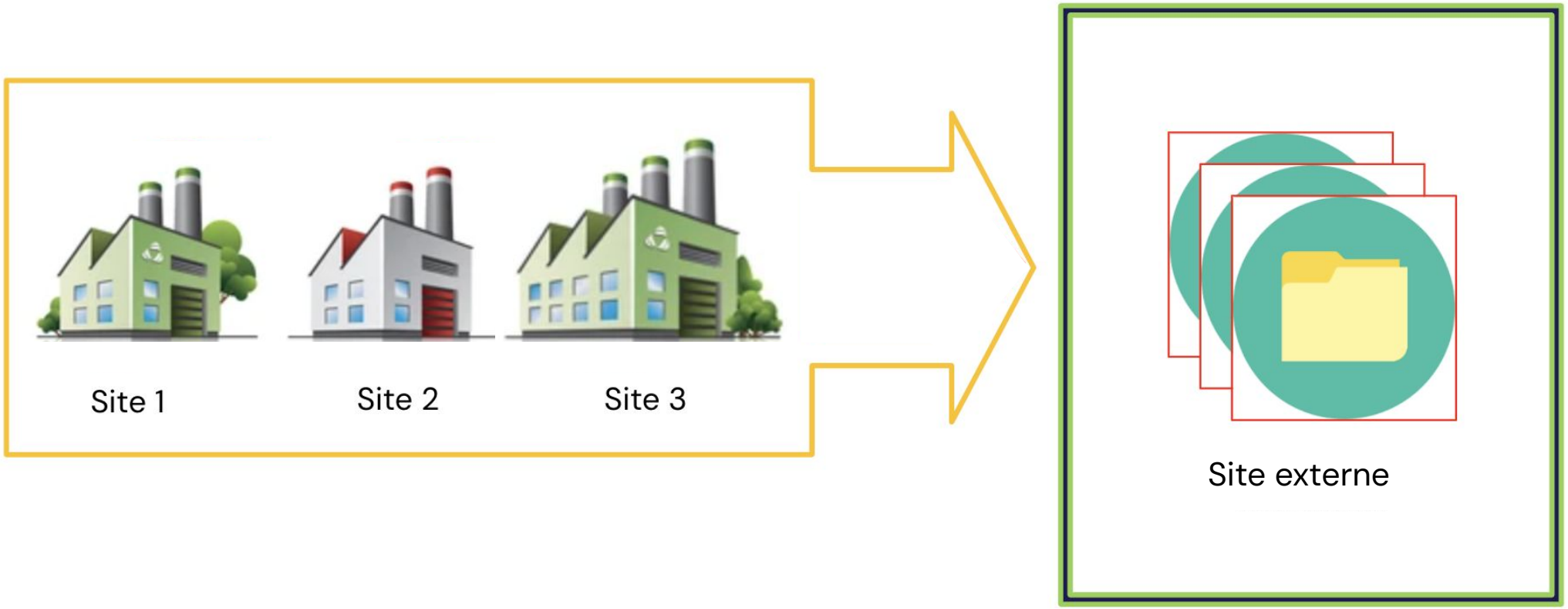
Apprentissage traditionnel | Dataset



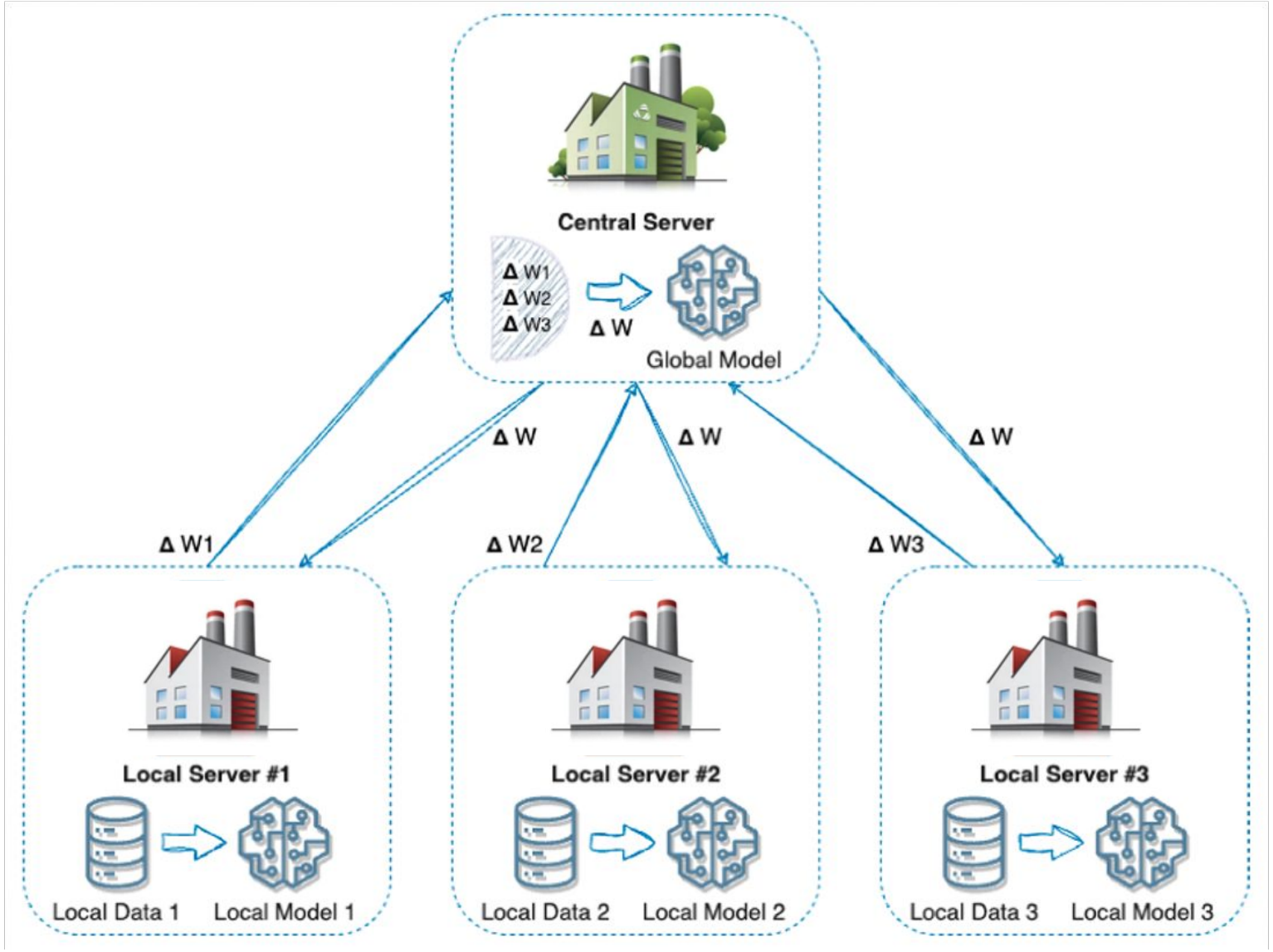
Apprentissage traditionnel | Modèle prédictif de maintenance



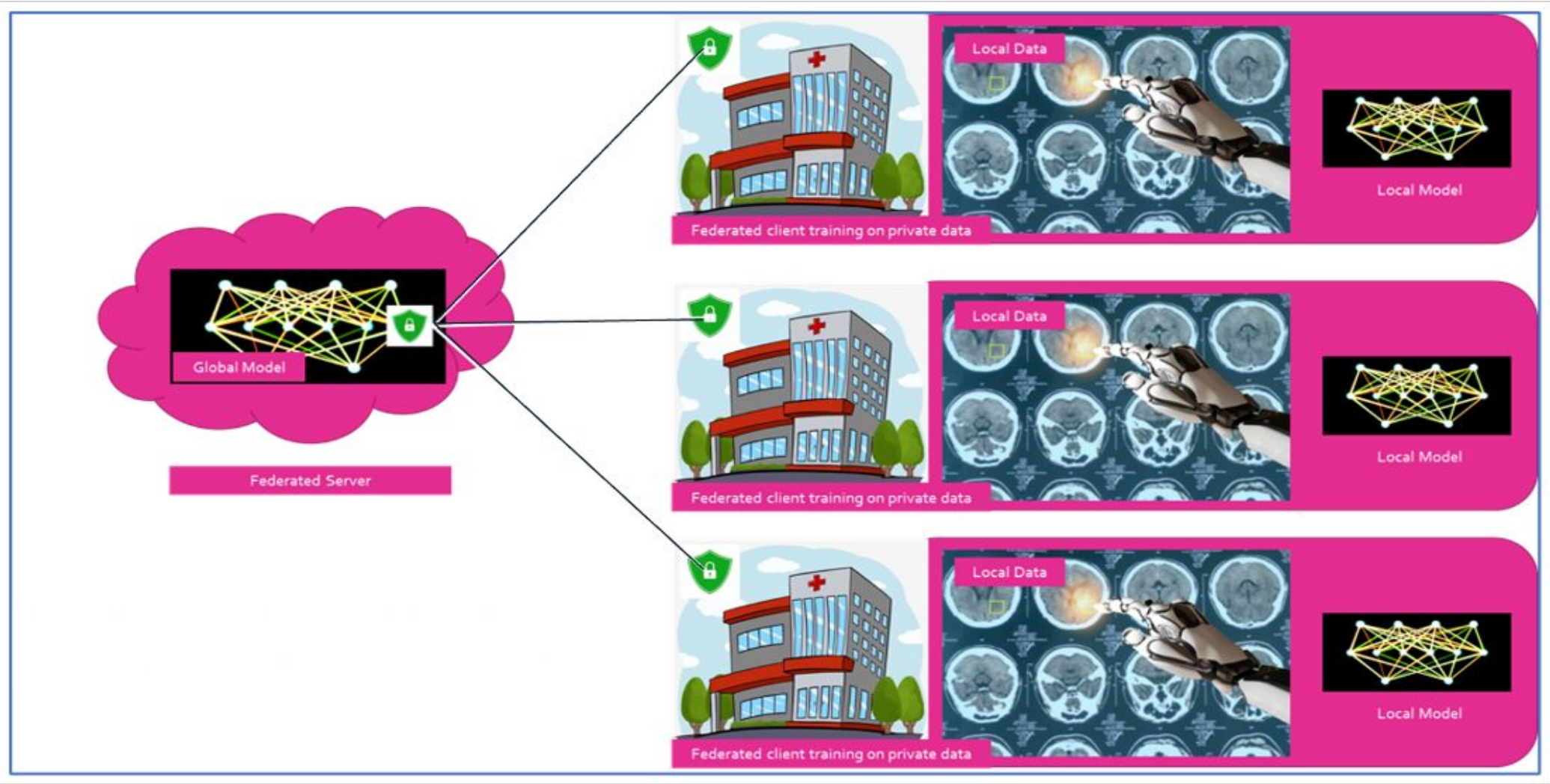
Apprentissage traditionnel | Dataset sur plusieurs sites



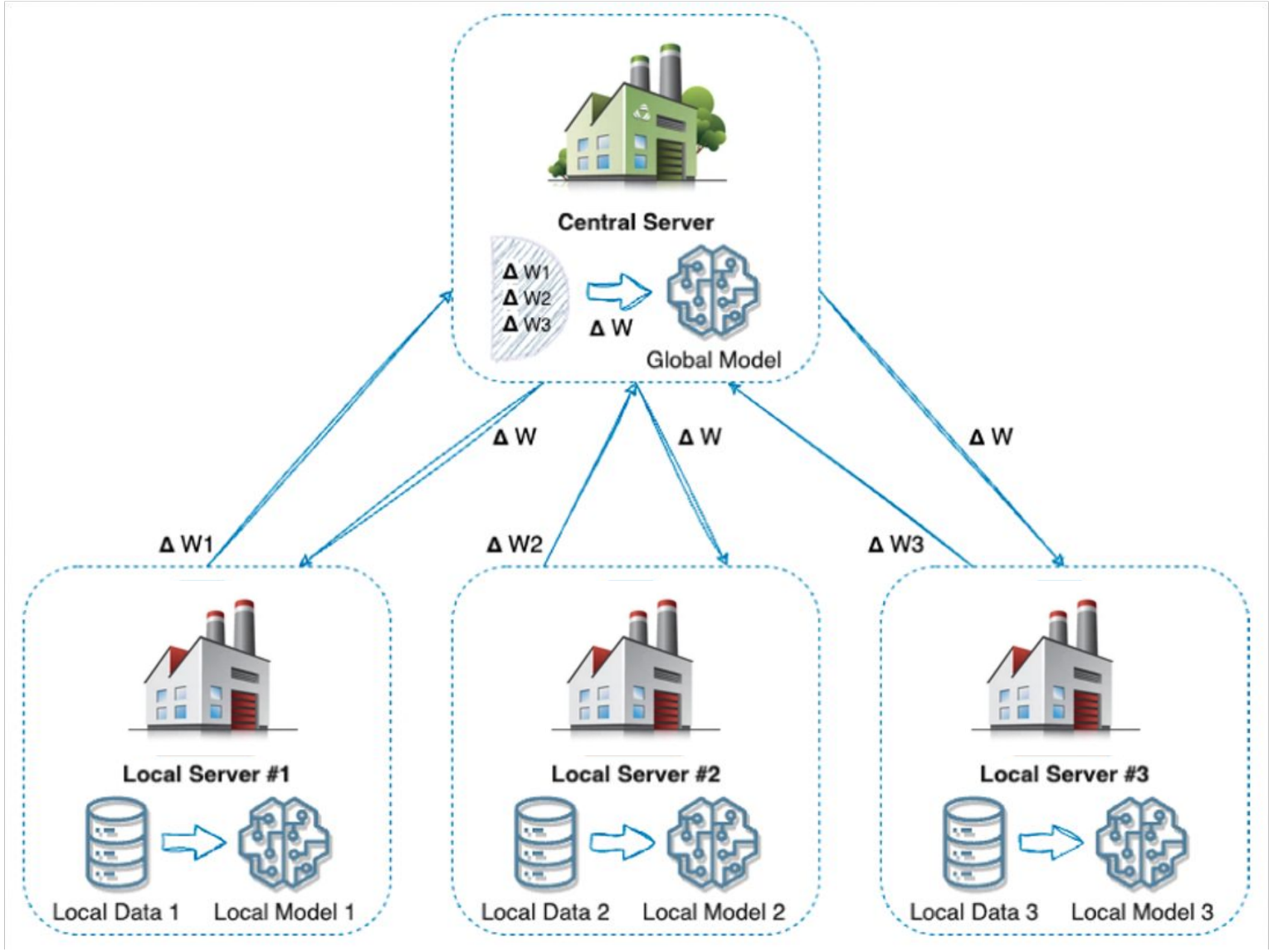
Apprentissage fédéré



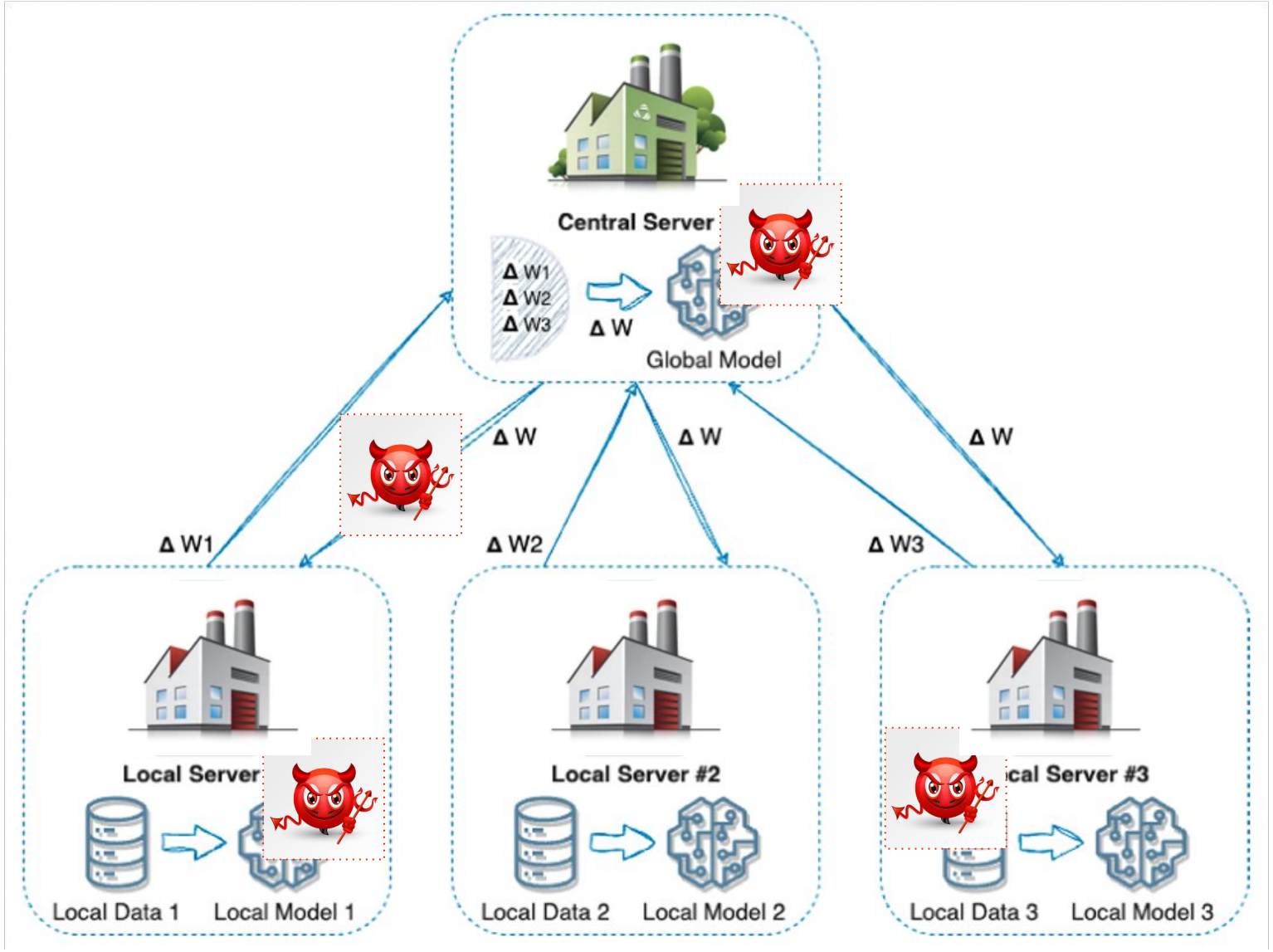
Apprentissage fédéré



Apprentissage fédéré



Risques de sécurité liés à l'apprentissage fédéré



Risques de sécurité liés à l'apprentissage fédéré

Attaques par empoisonnement (Poisoning Attacks)

Ces attaques consistent à injecter des données malveillantes ou à manipuler les mises à jour du modèle afin d'en dégrader volontairement les performances.

Attaques par porte dérobée (Backdoor Attacks)

Elles reposent sur l'entraînement de modèles locaux intégrant un déclencheur caché. Le modèle global se comporte normalement sur la majorité des entrées, mais produit systématiquement une mauvaise classification lorsqu'une entrée contient ce déclencheur spécifique.

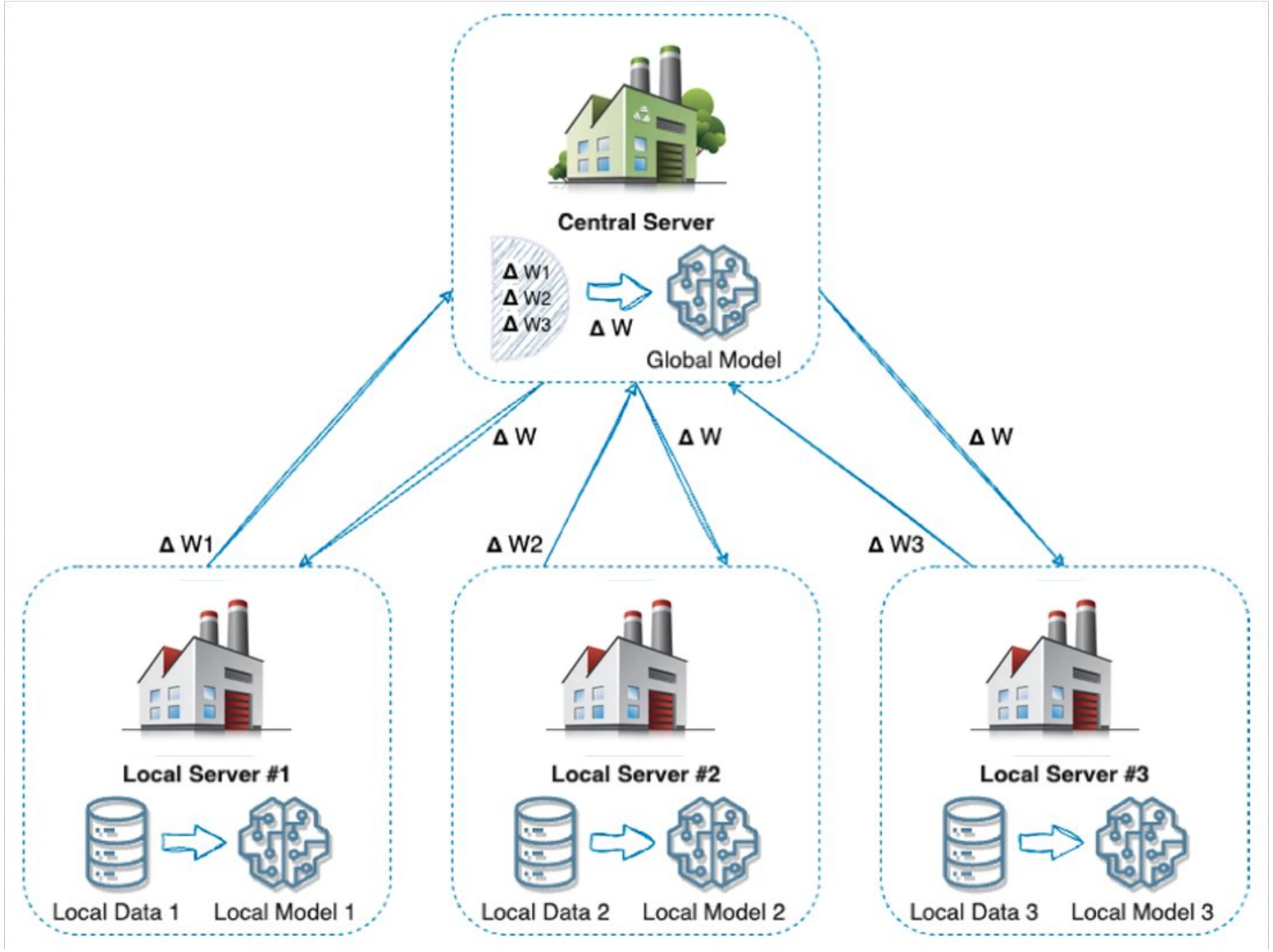
Attaques d'inférence (Inference Attacks)

Ces attaques visent à extraire des informations sensibles concernant les données d'entraînement à partir du comportement du modèle.

- **Attaques d'inversion de modèle (Model Inversion Attacks)** :
  Elles permettent de reconstruire partiellement ou totalement les données d'entrée à partir des sorties du modèle.

- **Attaques d'inférence de propriétés (Property Inference Attacks)** :
  Elles cherchent à déduire des propriétés globales du jeu d'entraînement (ex. distributions démographiques, présence d'une classe spécifique).
- …

Apprentissage fédéré sécurisé

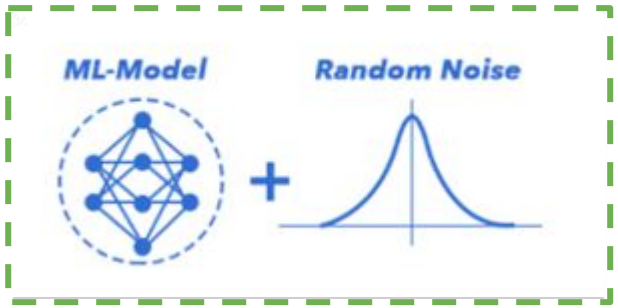


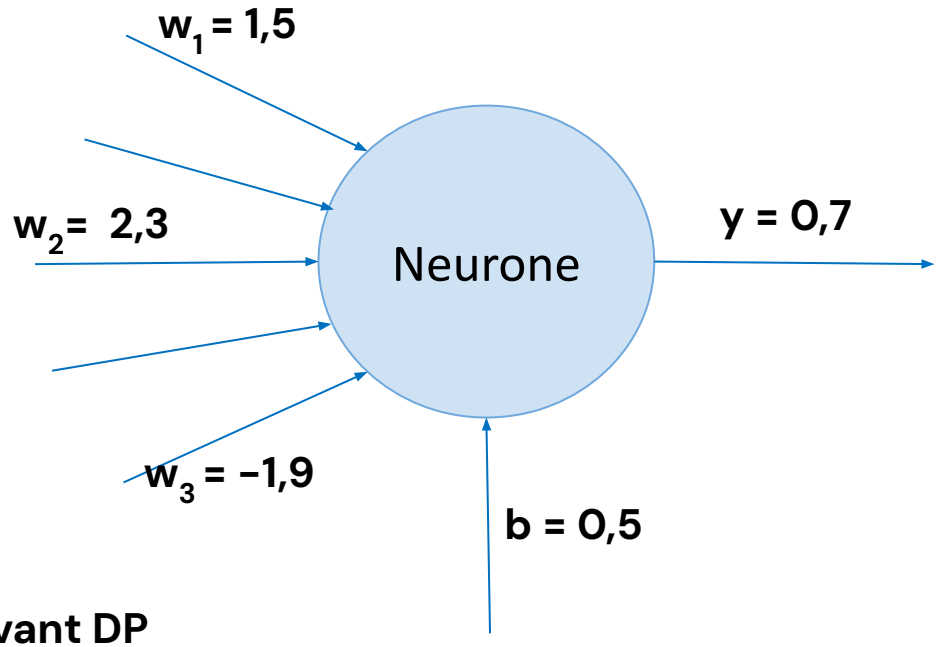
Architecture d'apprentissage fédéré intégrant

- Differential Privacy (DP)
- Fully Homomorphic Encryption (FHE)

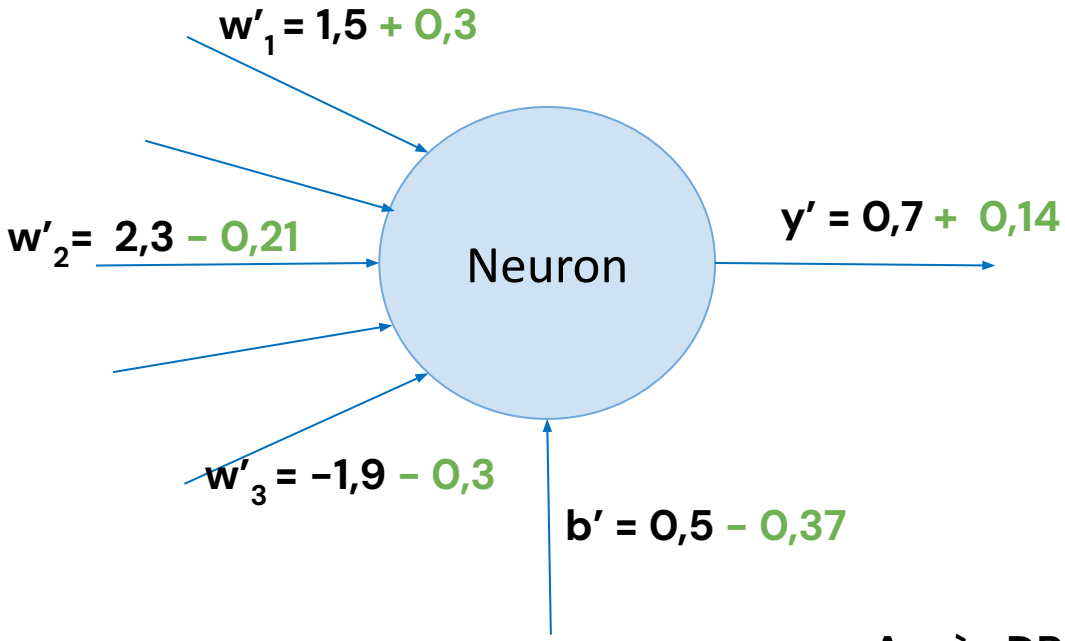
Confidentialité différentielle appliquée au modèle

Bruit Gaussien



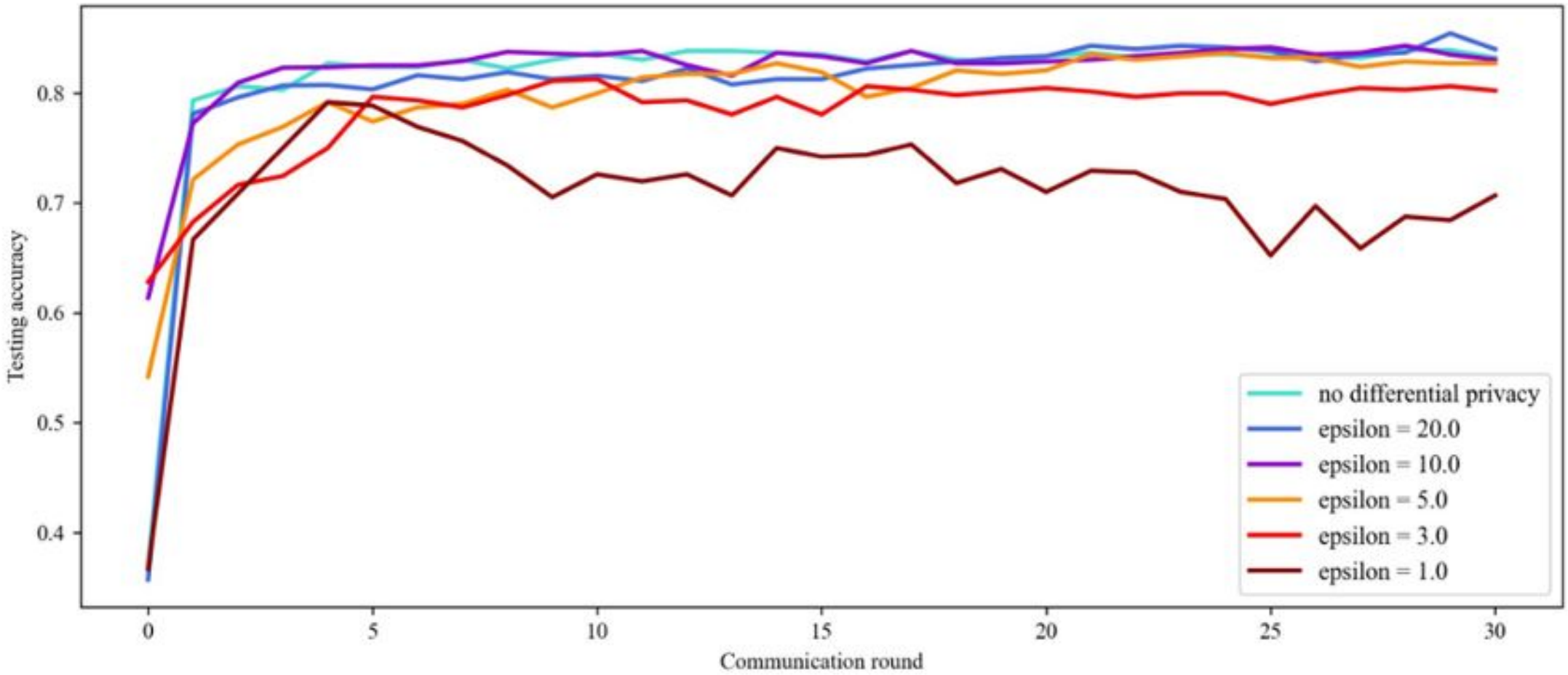


Avant DP

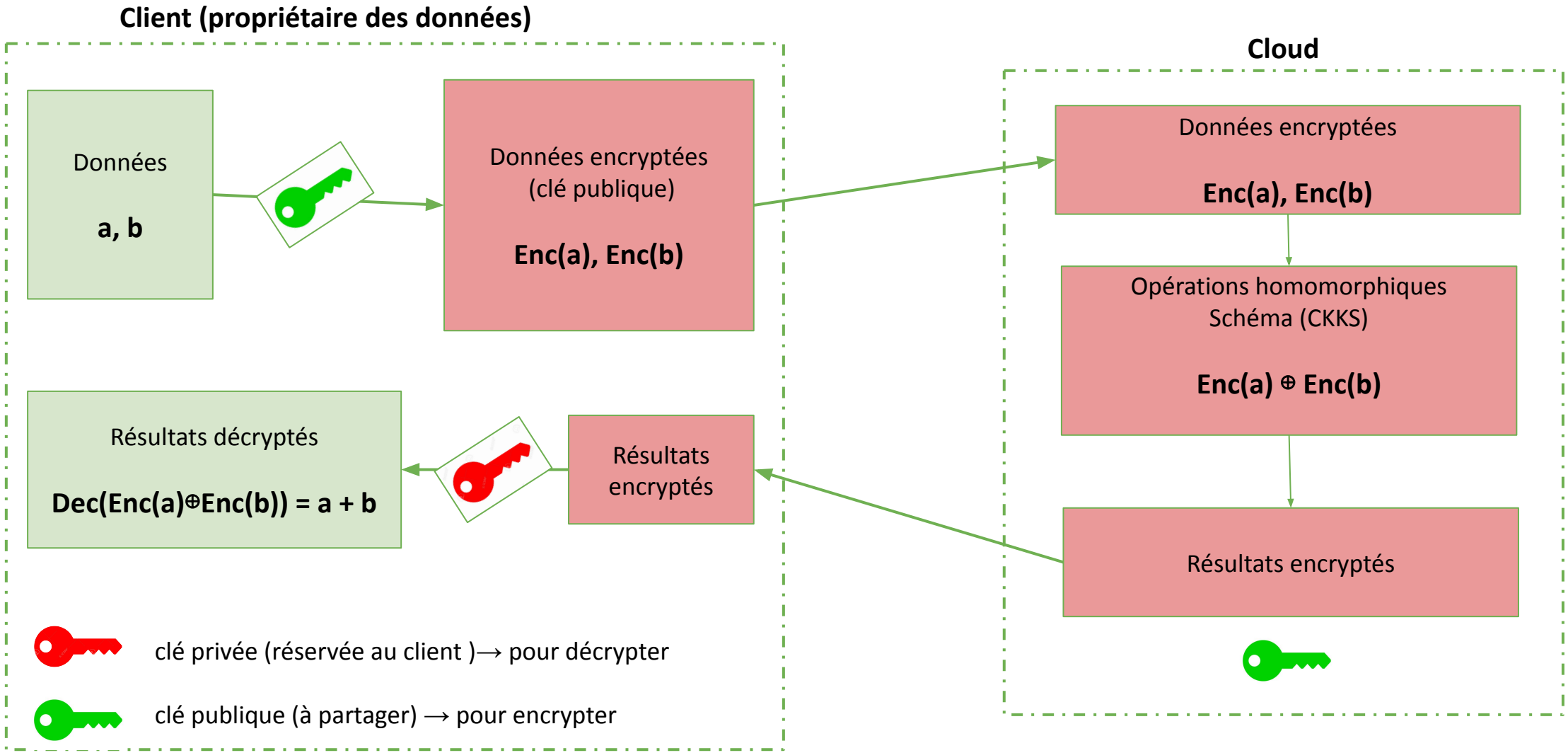


Après DP

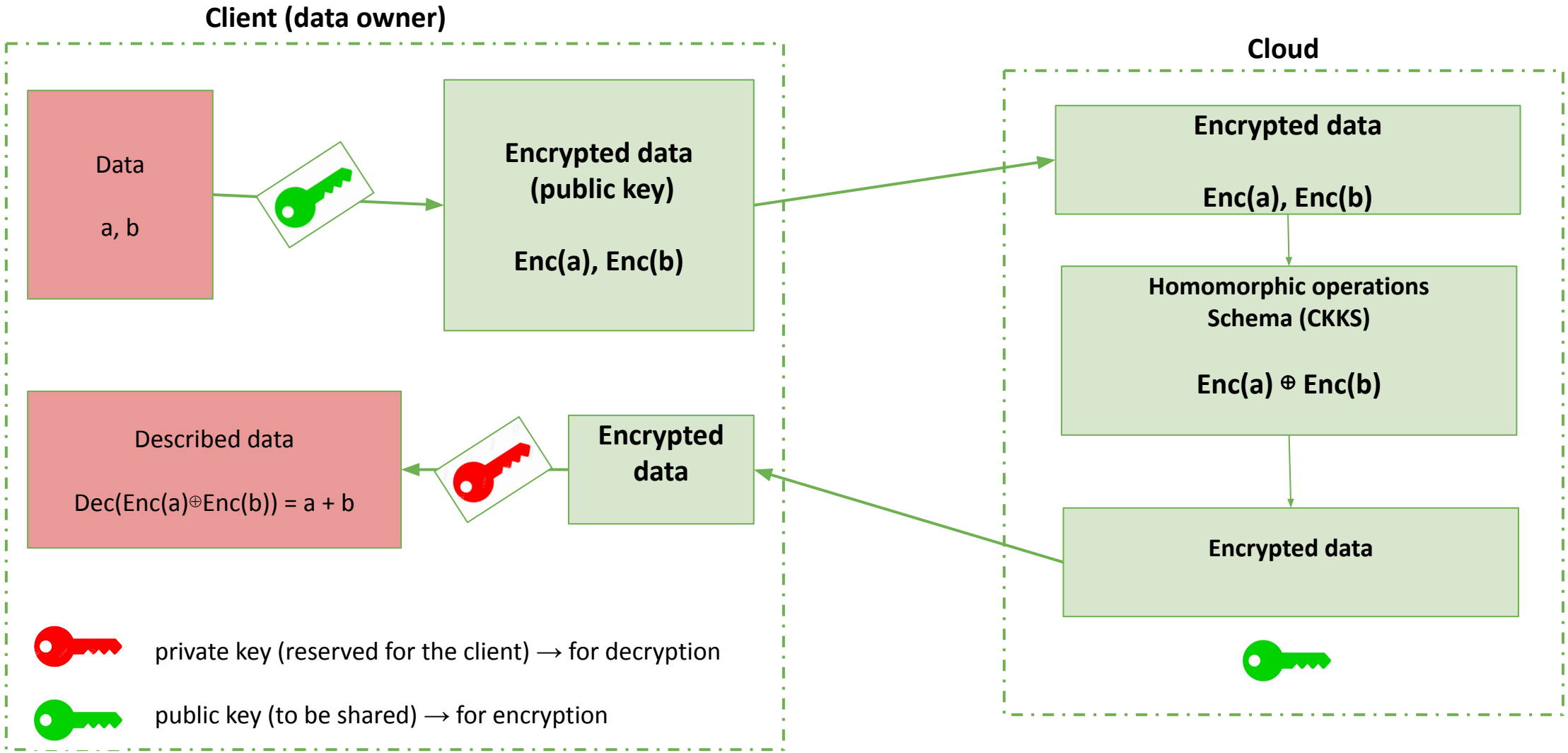
Confidentialité différentielle appliquée au modèle



Chiffrement homomorphe



Homomorphic encryption



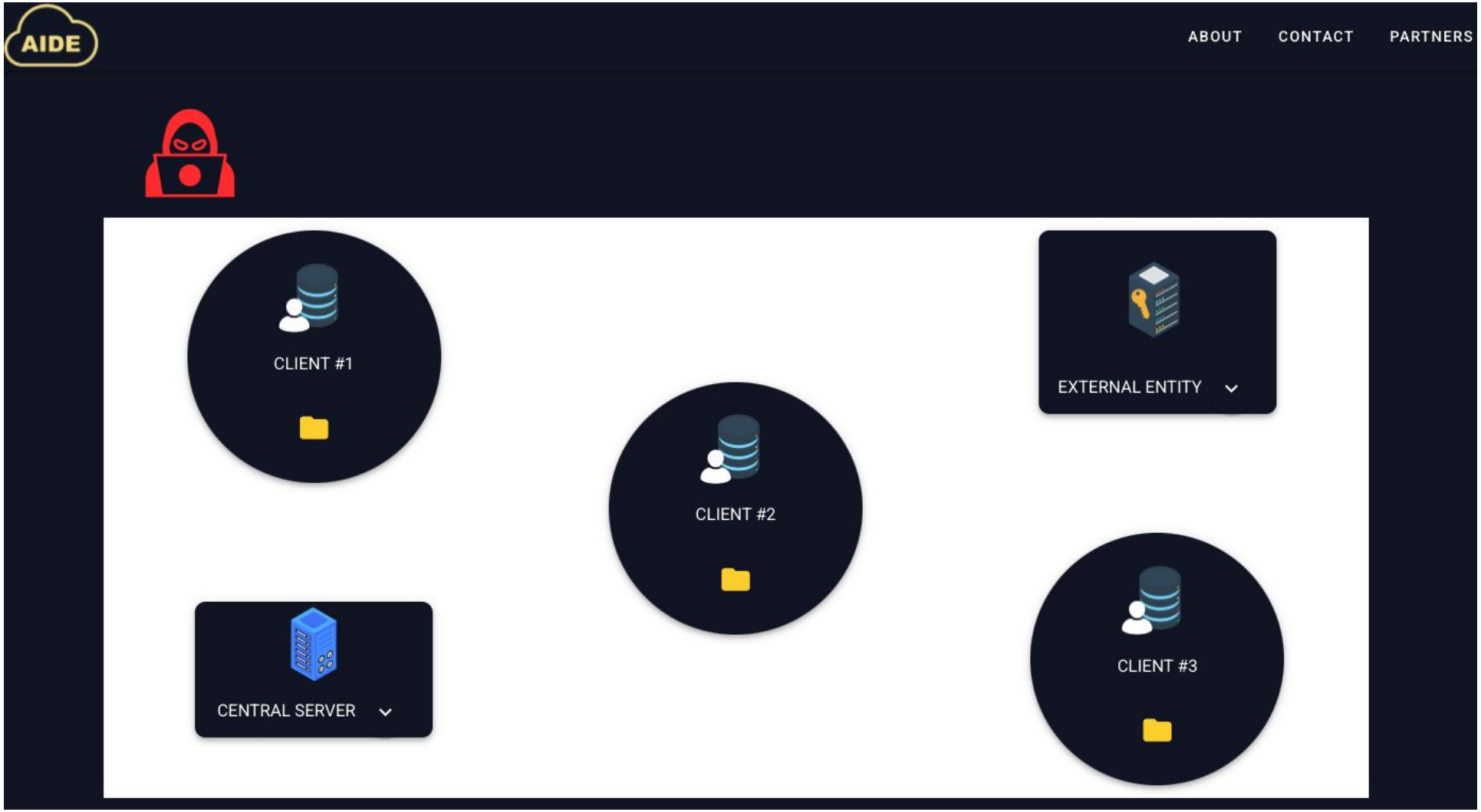
Chiffrement homomorphe appliqué au modèle | Poisoning

```
Before modification:
conv1 weight sample: tensor([[ 0.0838,  0.0946, -0.1161, -0.2028,  0.1563],
        [ 0.0607, -0.0285,  0.0219, -0.0759, -0.1135],
        [-0.0189, -0.1324, -0.0217,  0.1075, -0.0440],
        [ 0.0821, -0.0778, -0.0607,  0.0621, -0.0551],
        [-0.0650,  0.0537, -0.1490, -0.0754,  0.1101]],
       grad_fn=<SliceBackward0>)
Modifying: conv1.weight | Noise std: 0.99
Modifying: conv2.weight | Noise std: 0.99
Modifying: fc1.weight | Noise std: 0.99
Modifying: fc2.weight | Noise std: 0.99
Modifying: fc3.weight | Noise std: 0.99

After modification:
conv1 weight sample: tensor([[-0.2021,  0.3936, -0.5361, -0.6707,  1.5427],
        [ 0.1703, -2.6256,  2.2687, -2.3186,  0.0254],
        [ 1.1840,  0.8813,  0.2577,  0.9773, -1.3906],
        [ 1.1486,  1.9289, -1.5879,  2.1437,  0.8470],
        [ 0.7635,  1.1911,  0.6292,  0.8935, -0.2400]],
       grad_fn=<SliceBackward0>)
Modèle modifié sauvegardé dans : /model/cifar_client_1.pt
```

```
Error executing poisoning script: Command '['python3', 'poisoning.py', '--model_path', '/model/server_client_1.pkl',
returned non-zero exit status 1.
```

Demonstrateur |Front end



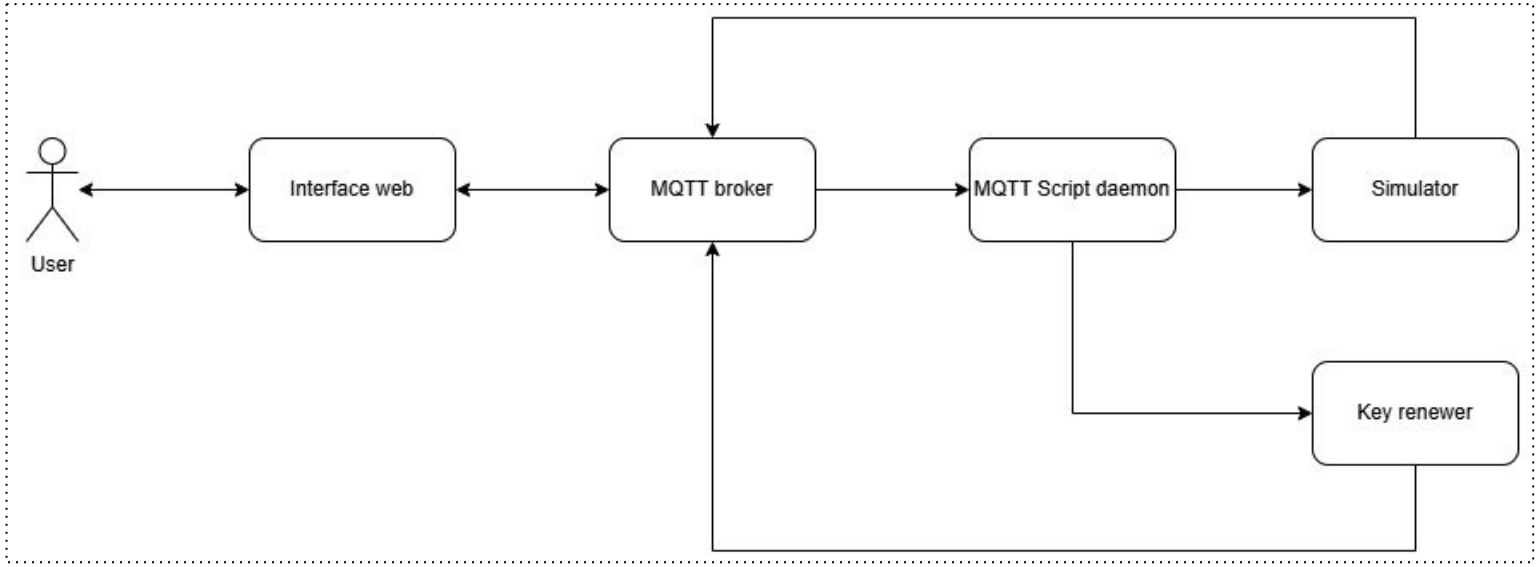
http://demo-aide.int.cetic.be

Démonstrateur | Back end

- 3 nodes / CIFAR-10
- Secure federated learning
  - DP / Opacus
  - FHE / CKKS

Flower

- Pods / Persistent volumes

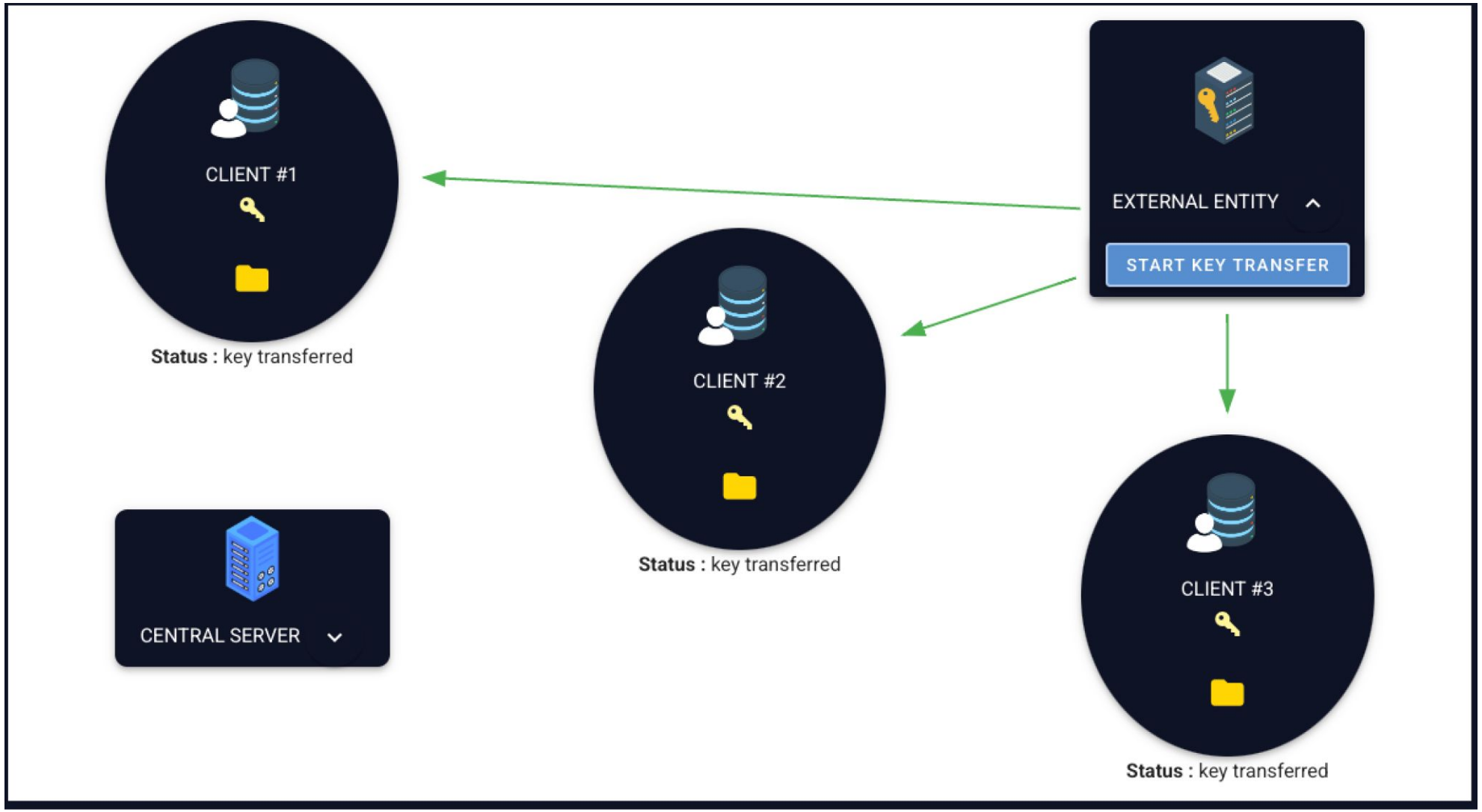



```
xle@node1:~$ sudo kubectl get pods
NAME                                READY   STATUS    RESTARTS      AGE
aide-client-0-58f475dfc-pf4jc       1/1     Running   0             19h
aide-client-1-56955f7cdd-bqx9h      1/1     Running   0             20h
aide-client-2-7cb554fdd7-9n7ff      1/1     Running   0             20h
aide-key-renewer-77d8fcb8f6-r6jhb   1/1     Running   7 (14d ago)   98d
aide-server-5d796f9c85-rqh5w        1/1     Running   0             20h
emqx-5f979cfc9c-mlw2l               1/1     Running   6 (14d ago)   94d
xle@node1:~$ sudo kubectl get pvc
NAME                  STATUS   VOLUME                  CAPACITY   ACCESS MODES   STORAGECLASS   VOLUMEATTRIBUTESCLASS   AGE
nfs-client0-data-pvc  Bound    nfs-client-0-data-pv    100Gi      RWX                           <unset>                 135d
nfs-client1-data-pvc  Bound    nfs-client-1-data-pv    100Gi      RWX                           <unset>                 135d
nfs-client2-data-pvc  Bound    nfs-client-2-data-pv    100Gi      RWX                           <unset>                 135d
nfs-clients-keys-pvc  Bound    nfs-clients-keys-pv     200Mi      RWX                           <unset>                 136d
nfs-model-pvc         Bound    nfs-model-pv            100Gi      RWX                           <unset>                 92d
nfs-server-data-pvc   Bound    nfs-server-data-pv      100Gi      RWX                           <unset>                 136d
nfs-server-key-pvc    Bound    nfs-server-key-pv       200Mi      RWX                           <unset>                 136d
```

MQTT

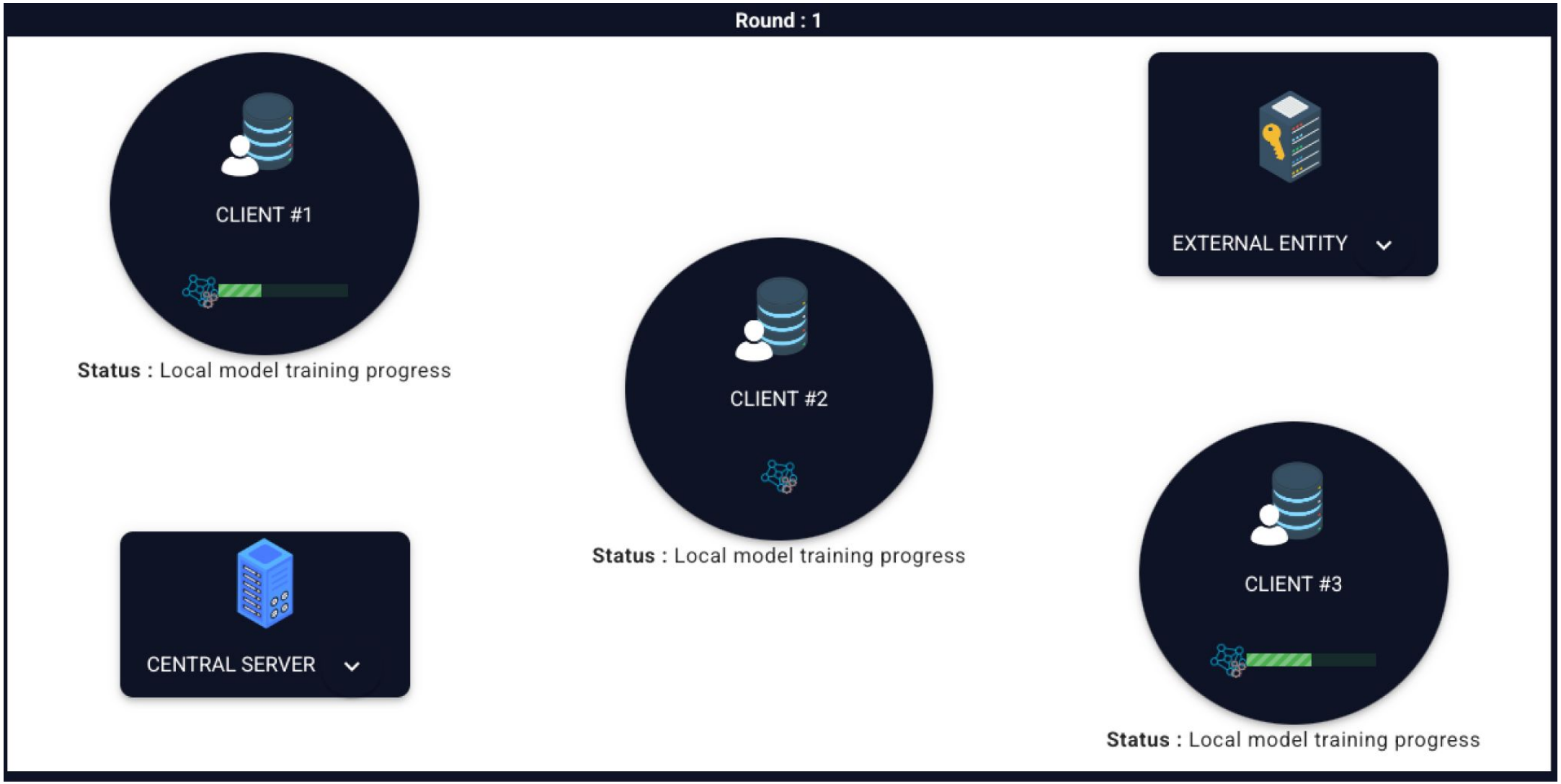
kubernetes

Démonstrateur / Screenshot



http://demo-aide.int.cetic.be

Démonstrateur / Screenshot



http://demo-aide.int.cetic.be

Questions ?

Merci pour votre attention

Apprentissage fédéré au service d'une IA confidentielle

Xavier Lessage, Ir, PhD

Expert researcher (IA, Cyber) @ CETIC

xavier.lessage@cetic.be

https://www.linkedin.com/in/xavier-lessage/

cetic

Your Connection to ICT Research

Aéropole
Avenue Jean Mermoz 28
6041 Charleroi - Belgique

twitter.com/@CETIC
twitter.com/@CETIC_be

linkedin.com/company/cetic

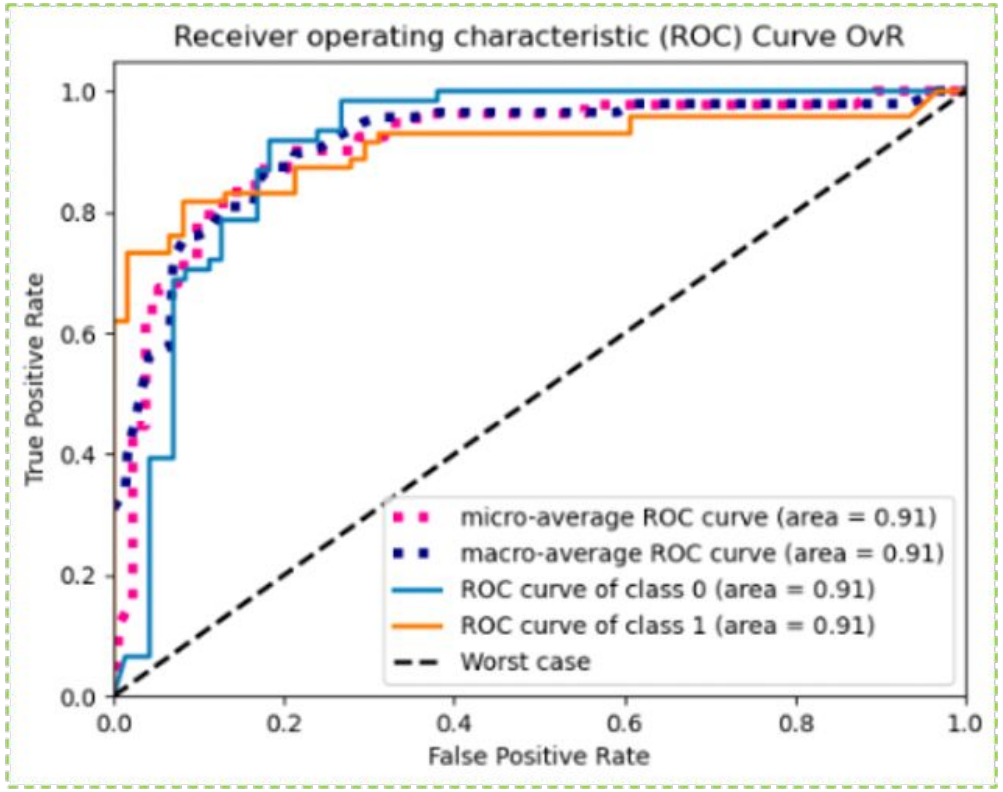
info@cetic.be

+32 71 159 362

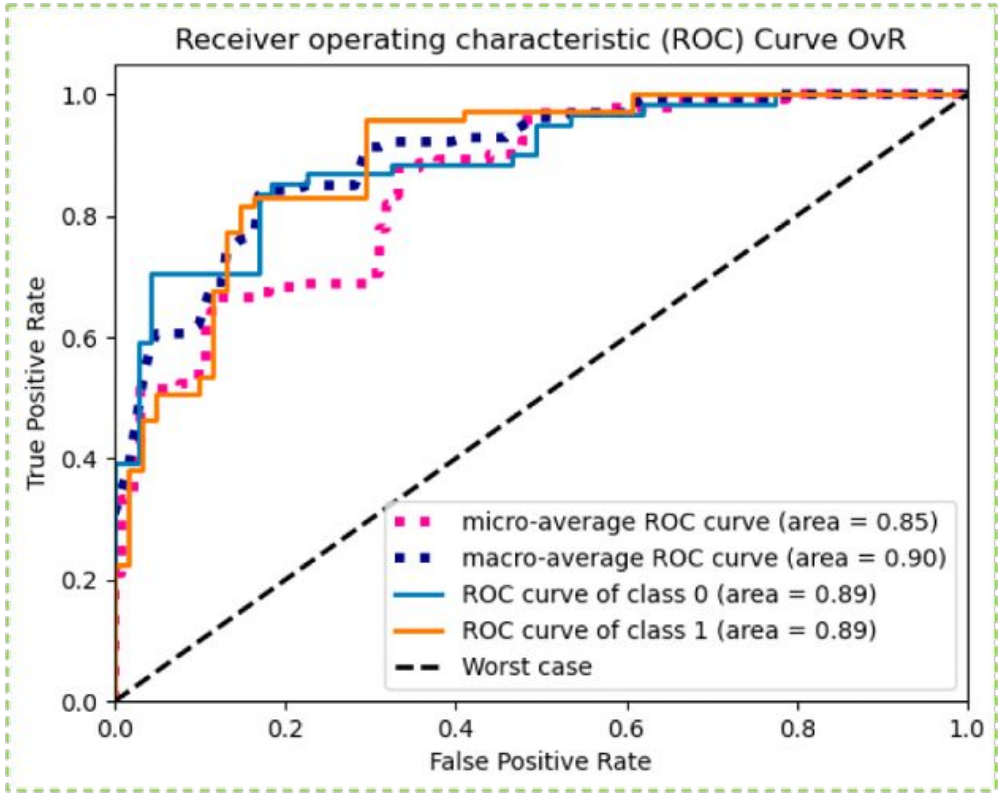
Videos

- FL + DP
- FL + FHE
- FL + FHE + Poisoning attack

Secure Federated learning with full homomorphic encryption



With FHE



Without FHE

Xavier Lessage, Leandro Collier, Saïd Mahmoudi, Axel Legay, Secure federated learning applied to medical imaging with fully homomorphic encryption, ICAIC'2024, IEEE Xplore.

Risques de sécurité liés à l'apprentissage fédéré

Defensive Techniques	Attack Types			
	Model Poisoning	Data Poisoning	Backdoor Attacks	Privacy Inference
Anomaly Detection (statistical analysis or performance metrics)	YES	YES	YES	NO
Differential Privacy (local or central)	NO	NO	YES	YES
Homomorphic Encryption (HE)	NO	NO	NO	YES
Secure Multi-party Computation (SMPC)	NO	NO	NO	YES
Encrypted Communication (TLS)	NO	NO	NO	YES