# Federated Learning Deep-dive

Inpyo-Hong

# Introduction

- **Why AI application is necessary in the medical field?**
  - 의료진의 업무 경감
  - 진단 보조기구의 활용을 통한 신속하고 정확한 진료
  - 의료인력 절감을 통한 의료비 경감

- **Limitations of Medical Dataset**
  - 데이터 확보의 어려움 (ex. 환자의 소극적인 데이터 제공, 의료진의 라벨링 작업)
  - 보안적 측면의 위험성 (데이터 유출 시 민감한 개인정보 유출)

Reference: Poudel, Sahadev, et al. "Colorectal disease classification using efficiently scaled dilation in convolutional neural network." IEEE Access 8 (2020): 99227-99238.

# Introduction

- **Key points for applying AI to the medical field**

① **데이터 확보**

- **안전한 AI학습** 방법 ➔ 데이터 제공할 환자에게 신뢰감 형성 ➔ **연합학습 적합**

② **일반화 성능 향상**

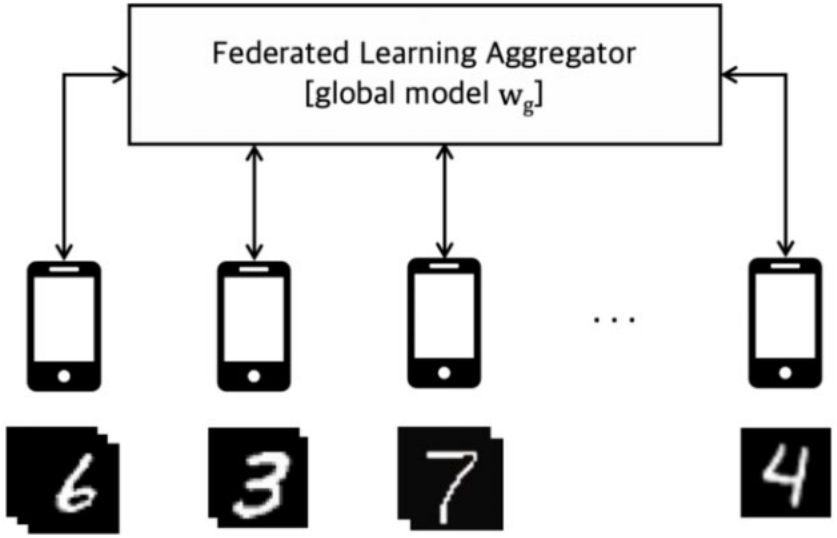
- 같은 종류의 데이터라도 의료장비의 종류에 따라 학습결과는 천차만별 (Multi-modal dataset)

➔ **일반화 성능을 향상한 AI 학습** 필요 ➔ **연합학습 부적합 (∵ Non-IID Dataset)**

Reference: Castellanos, Julio, Gonzalo Raposo, and Lucia Antunez. "Data Federation in Healthcare for Artificial Intelligence Solutions." Stud. Health Technol. Inform 295 (2022): 167-170.

# Limitation of Federated learning

- **Limitation of Federated learning for medical field**
  - Non-IID상황에서의 글로벌 최적화 모델 생성 어려움




- **Non-IID 데이터란?**

  Client가 소유한 각 데이터가 독립되어 있고,
  그 데이터가 동일한 확률분포 X

Ex.

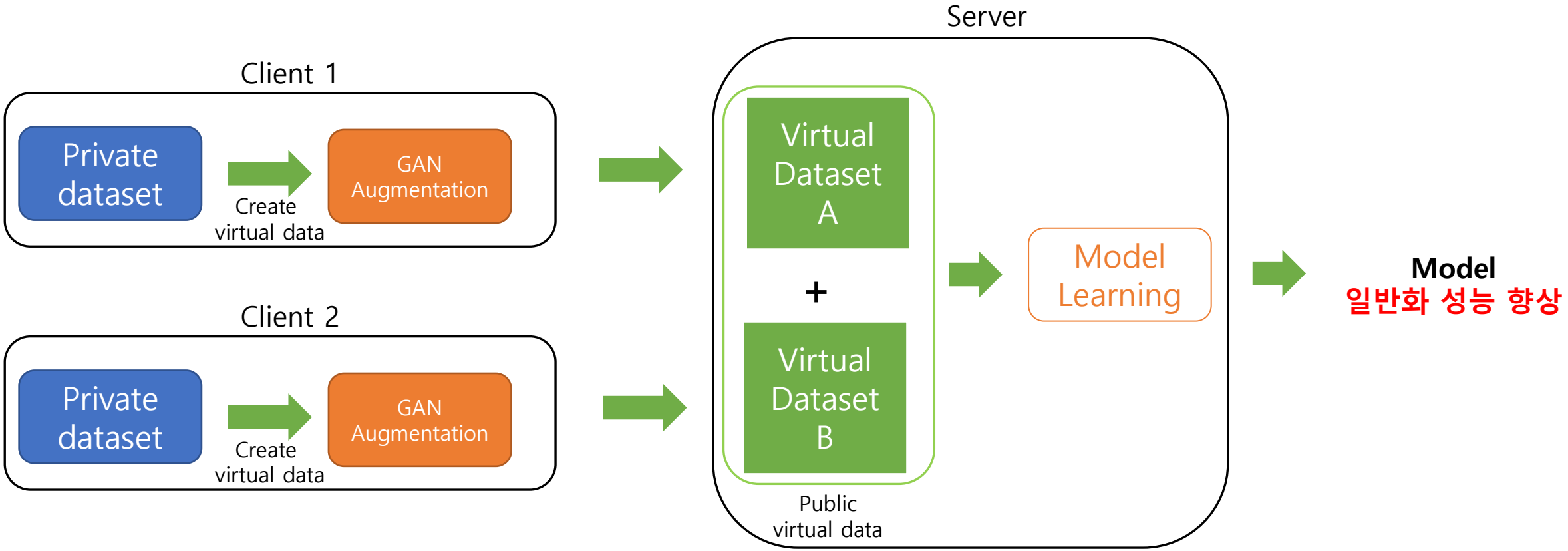
한국인의 이미지데이터셋: 동양인 사진 ↑
미국인의 이미지데이터셋: 서양인 사진 ↑
∴ **FL 학습시 나라간 로컬 업데이트 차이 ↑**

**Thus, Global한 학습 어려움**

Reference: Kairouz, Peter, et al. "Advances and open problems in federated learning." *Foundations and Trends® in Machine Learning* 14.1–2 (2021): 1-210.

# Methods

- **Creating a virtual public dataset through Generative Adversarial Network (GAN)**



# Methods

- **Generative Adversarial Network (GAN)**

이 사람들은 존재하지 않습니다. 모두 GAN이 만든 가상의 인물입니다.

  - **가상의 이미지를 생성하여 개인정보 유출 우려 X**
  - **Client가 보유한 데이터셋의 특성에 따른 데이터셋 생성**

# Conclusion

- Client상에서 Generative Adversarial Network (GAN)기법을 활용한 가상 데이터셋 생성
- Server상에서 사용할 수 있는 공용 데이터셋 구성
- Server상에서 발생할 수 있는 일반화 성능 저하 문제를 데이터셋 유출 없이 해결