Federated Learning on Non-IID Data

Presenter:

Xin Fan

2021/09/22

Outline

- Problem Setup
- Main Approaches to Handling Non-IID Data
  - Data based approach
  - Algorithm based approach
  - System based approach
- Discussion

Outline

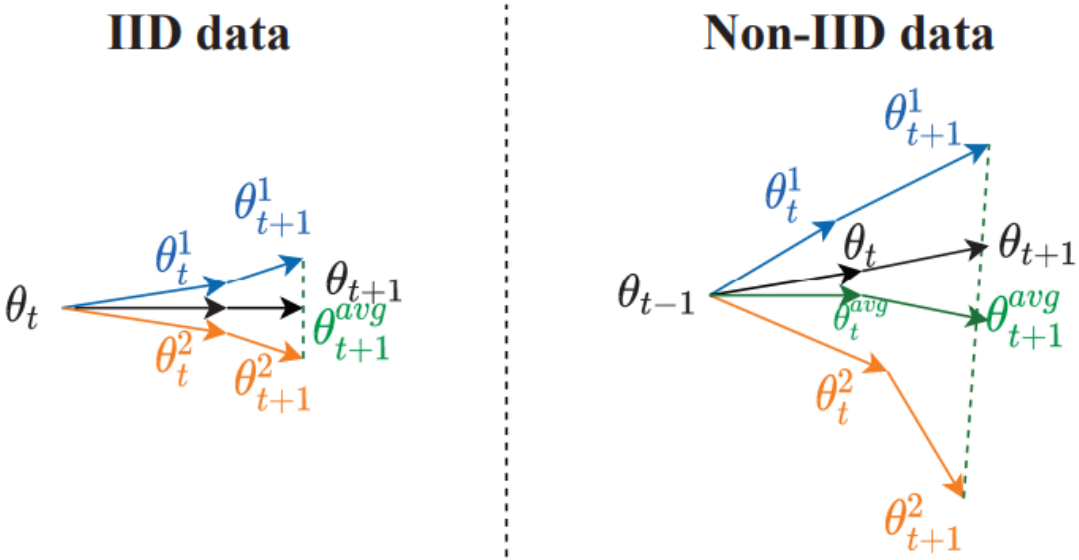
- **Problem Setup**
- Main Approaches to Handling Non-IID Data
  - Data based approach
  - Algorithm based approach
  - System based approach
- Discussion

Problem Setup

- FedAvg
- Non-IID federated learning

	bird	deer	frog	ship
Client 1				
Client 2				

- An illusion of parametric model divergence

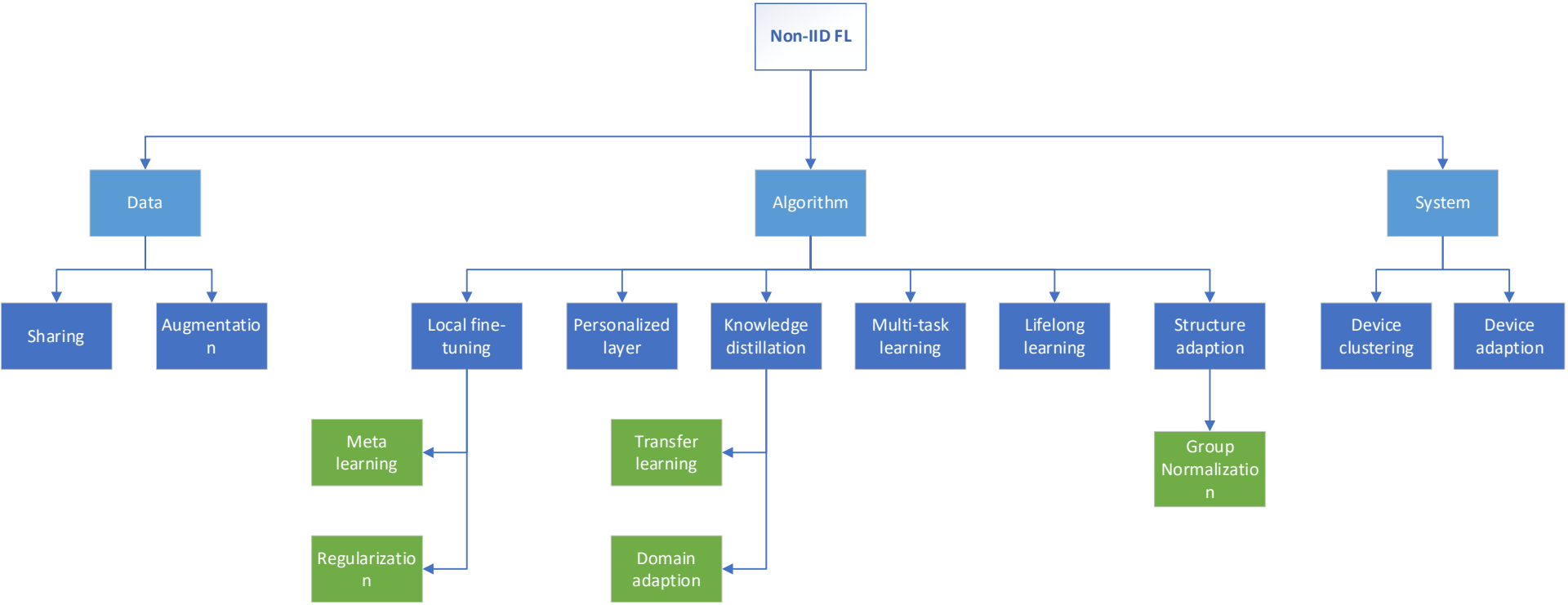


Outline

- Problem Setup
- **Main Approaches to Handling Non-IID Data**
  - Data based approach
  - Algorithm based approach
  - System based approach
- Discussion

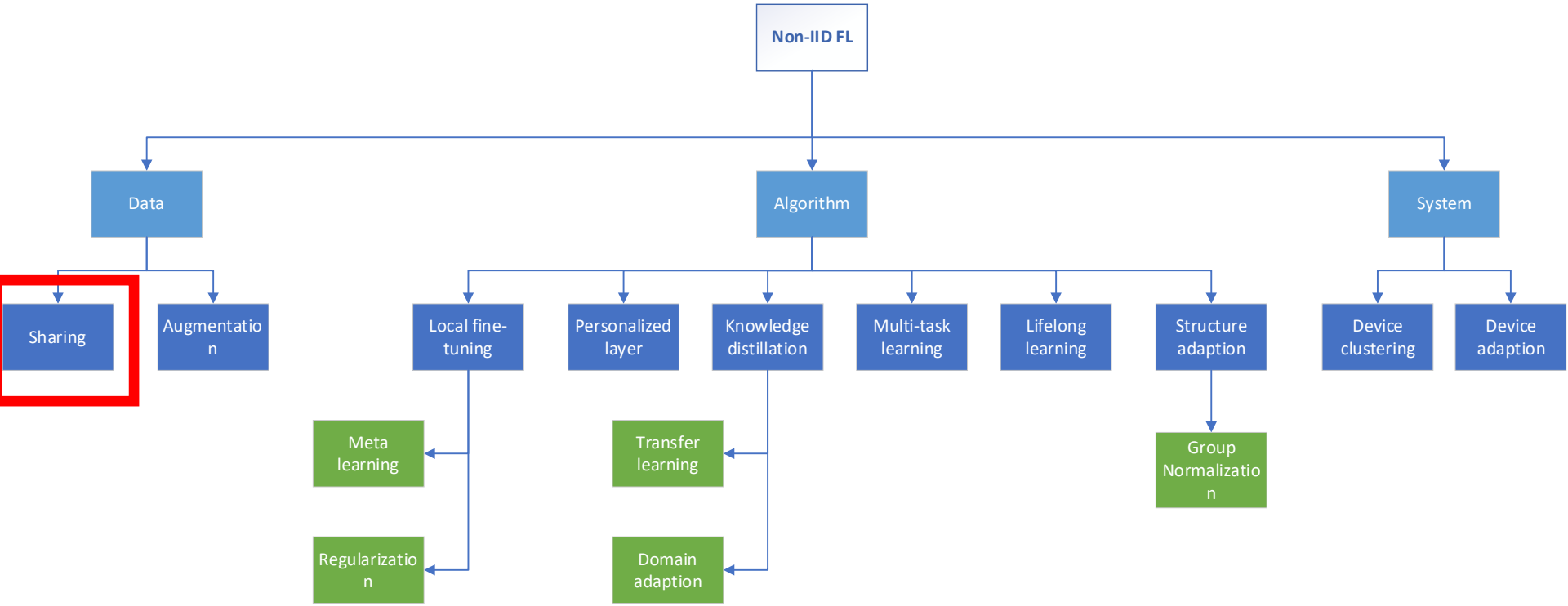
Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data



Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data

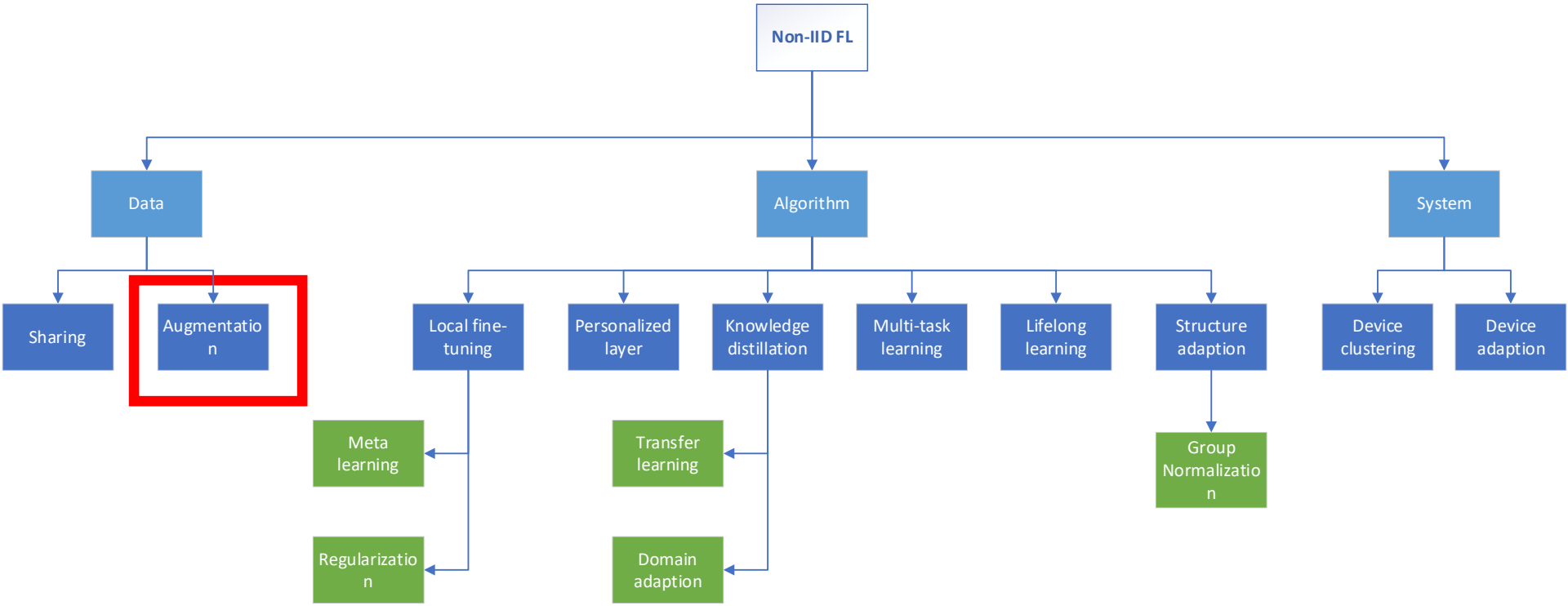


Main Approaches to Handling Non-IID Data

- Data sharing
  - [1] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., Chandra, V., 2018. Federated learning with non-iid data. arXiv preprint arXiv:1806.00582 .
    - Using a global dataset with a uniform distribution
    - Local models are updated by both training local data and the sharing global dataset
    - Shortcoming: it is hard to get so called uniformly distributed global dataset
  - [2] Tuor, T., Wang, S., Ko, B.J., Liu, C., Leung, K.K., 2020. Overcoming noisy and irrelevant data in federated learning. arXiv e-prints ,arXiv–2001.
    [3] Yoshida, N., Nishio, T., Morikura, M., Yamamoto, K., Yonetani, R., 2019. Hybrid-fl: Cooperative learning mechanism using non-iid data in wireless networks.
    - Sharing some local data with the server.
    - Shortcoming: privacy issue

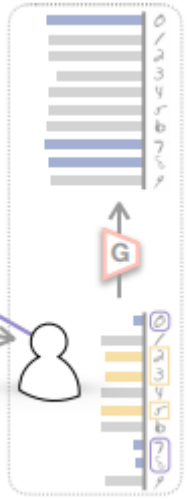
Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data



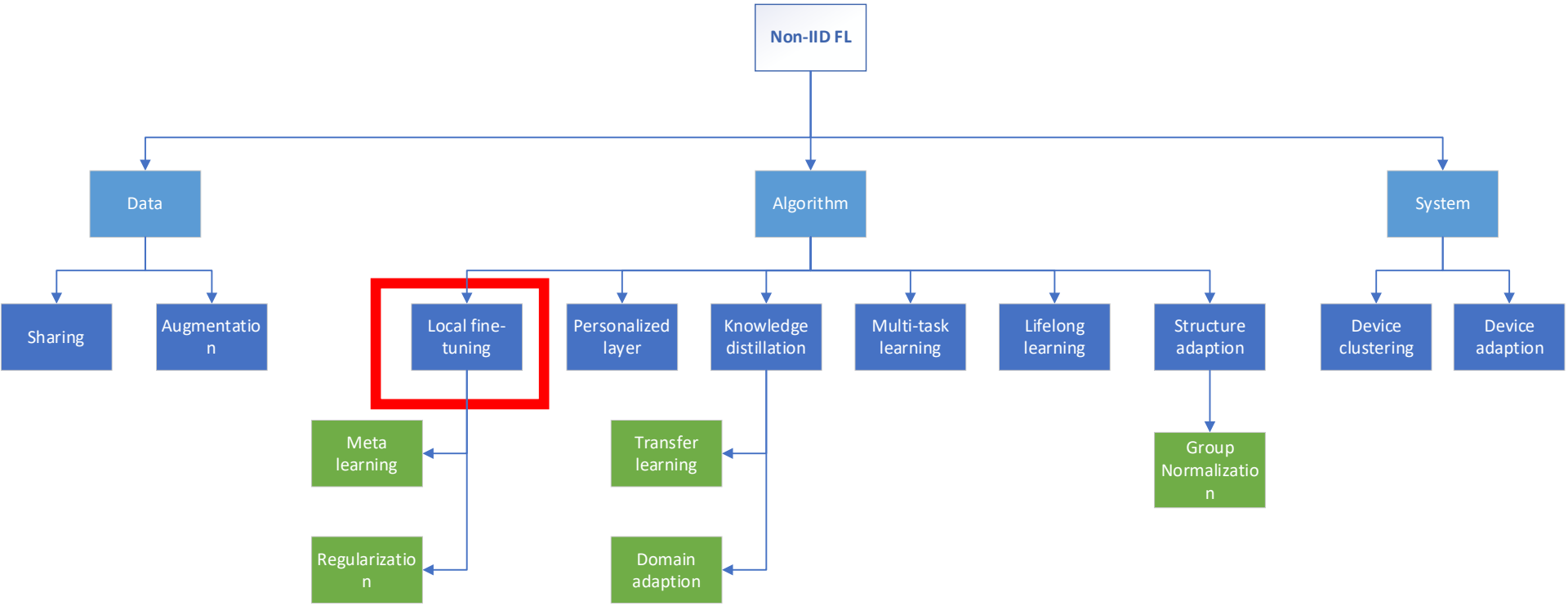
Main Approaches to Handling Non-IID Data

- Data Augmentation
  - [4] Duan, M., Liu, D., Chen, X., Tan, Y., Ren, J., Qiao, L., Liang, L., 2019. Astraea: Self-balancing federated learning for improving classification accuracy of mobile deep learning applications, in: 2019 IEEE 37th International Conference on Computer Design (ICCD), IEEE. pp. 246–254.
    - Displace some pixels of the original data samples to get new data samples
    - Both the original local data sample and the augmented data are used to update the local model parameters
  - [5] Jeong E, Oh S, Kim H, et al. Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data[J]. arXiv preprint arXiv:1811.11479, 2018.
    - The trained GAN empowers each device to locally reproduce the data samples of all devices, so as to make the training dataset become IID.



Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data



Main Approaches to Handling Non-IID Data

- Local fine-tuning
  - Meta learning
    - [6] Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach[J]. Advances in Neural Information Processing Systems, 2020, 33: 3557-3568.
      - Change the loss function

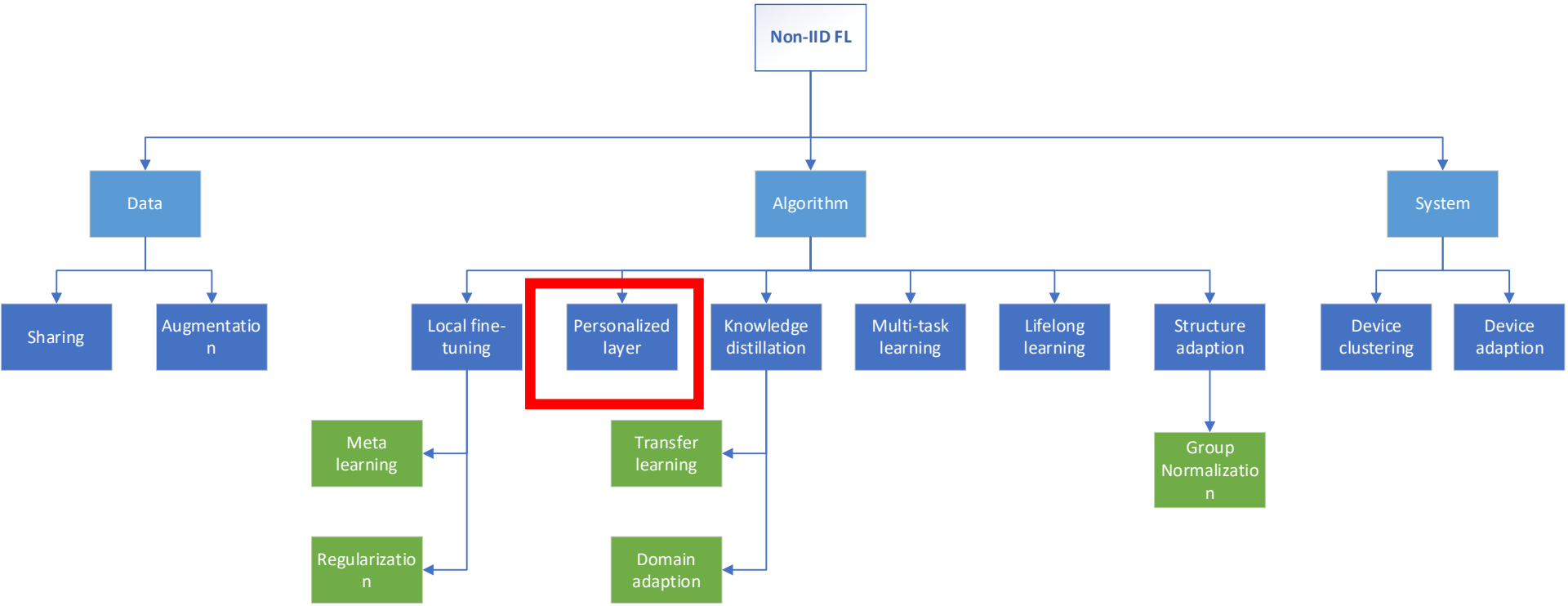
$$\min_{w\in\mathbb{R}^d} f(w) := \frac{1}{n}\sum_{i=1}^{n} f_i(w). \qquad \min_{w\in\mathbb{R}^d} F(w) := \frac{1}{n}\sum_{i=1}^{n} f_i(w-\alpha\nabla f_i(w)).$$

      - The meta-function is defined as $F_i(w) := f_i(w - \alpha\nabla f_i(w))$
  - Regularization
    - [7] Dinh, C.T., Tran, N.H., Nguyen, T.D., 2020. Personalized federated learning with moreau envelopes, in: Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual
    - Change the loss function, add a term of L2 norm

$$f_k(\theta_k) = \frac{1}{n_k}\sum_{i}^{n_k} l(\boldsymbol{x}_i, y_i; \theta_k) + \frac{\gamma}{2}||\theta_k - \theta||^2,$$

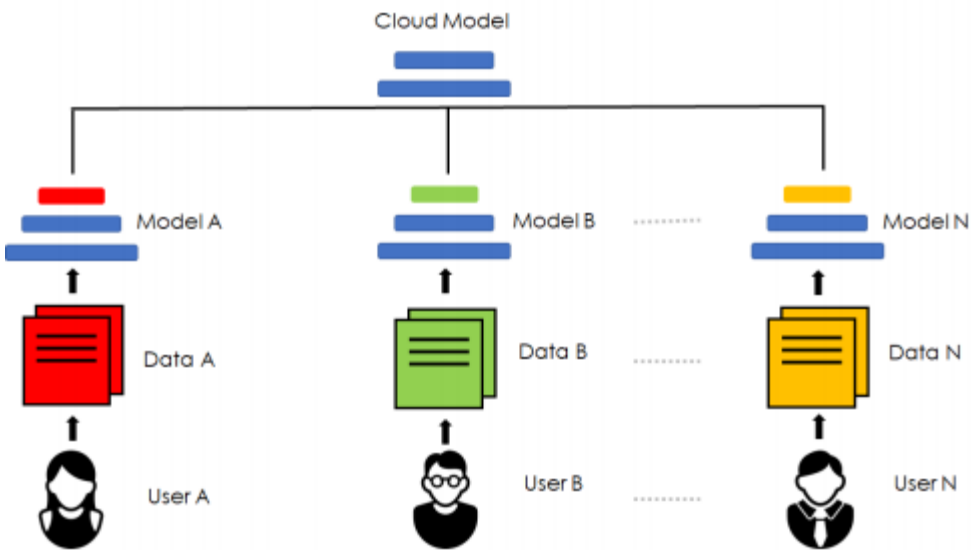
Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data



Main Approaches to Handling Non-IID Data

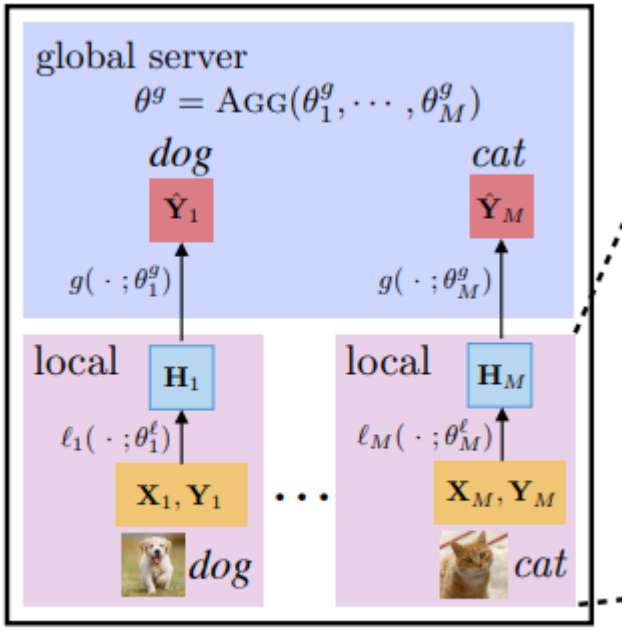
- Personalization layer
  - [8] Arivazhagan, M.G., Aggarwal, V., Singh, A.K., Choudhary, S., 2019. Federated learning with personalization layers. arXiv preprint arXiv:1912.00818 .
    - The base layers are the shallow layers of the neural network that extracts high-level representations and the personalization layers are the deep layers for classifications



Main Approaches to Handling Non-IID Data

- Personalization layer
  - [9] Liang, P.P., Liu, T., Ziyin, L., Allen, N.B., Auerbach, R.P., Brent, D., Salakhutdinov, R., Morency, L.P., 2020. Think locally, act globally: Federated learning with local and global representations. arXiv preprint arXiv:2001.01523
    - The personalization layers are shallow layers of the neural network and the base layers that are shared with the server are deep layers for class classifications.

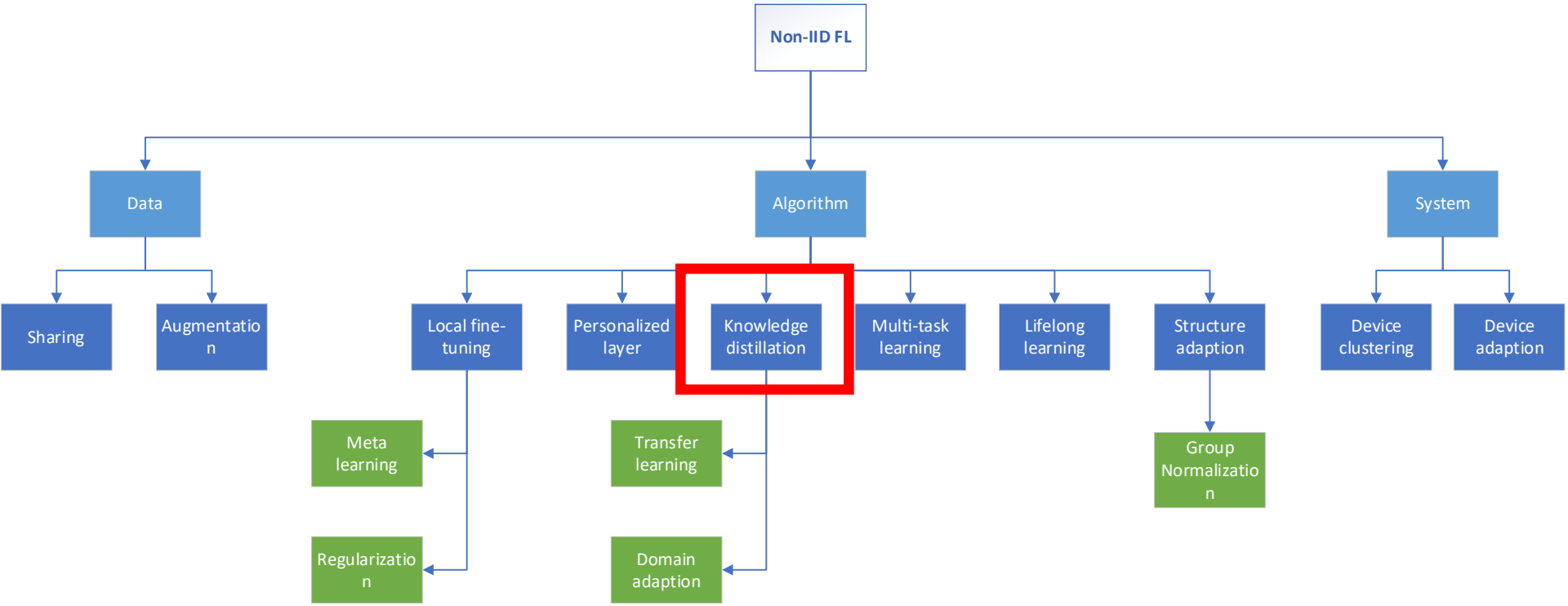
- Pros: reduce communication costs
- Cons: storage



Local Global Federated Averaging
(LG-FedAvg, our proposed method)

Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data



Main Approaches to Handling Non-IID Data

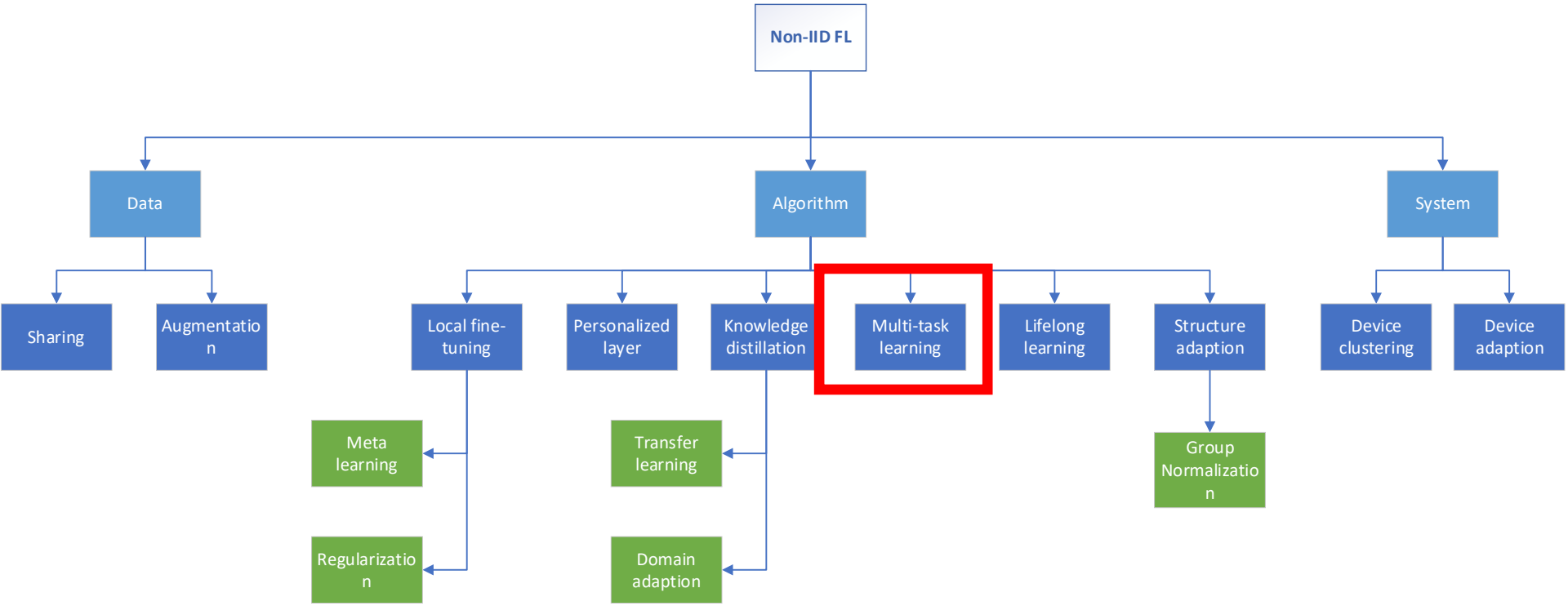
- Knowledge distillation
  - Transfer learning
    - Transfer knowledge from the server or other clients to a certain client to improve its performance on unknown heterogeneous data
    - [10] K. Wang, R. Mathews, C. Kiddon, H. Eichner, F. Beaufays, and D. Ramage, "Federated evaluation of on-device personalization," arXiv preprint arXiv:1910.10252, 2019.
      - Personalization on a well trained global model.
  - Domain adaption
    - [11] Peterson D, Kanani P, Marathe V J. Private federated learning with domain adaptation[J]. arXiv preprint arXiv:1912.06733, 2019.

$$\hat{y}_i = \alpha_i(x) M_G(x, \Theta_G) + (1 - \alpha_i(x)) M_{P_i}(x, \Theta_{P_i}).$$

$\alpha_i(x) = \sigma(w_i^T \cdot x + b_i)$, where $\sigma(x)$ is the sigmoid function, and w_i and b_i are learned weights.

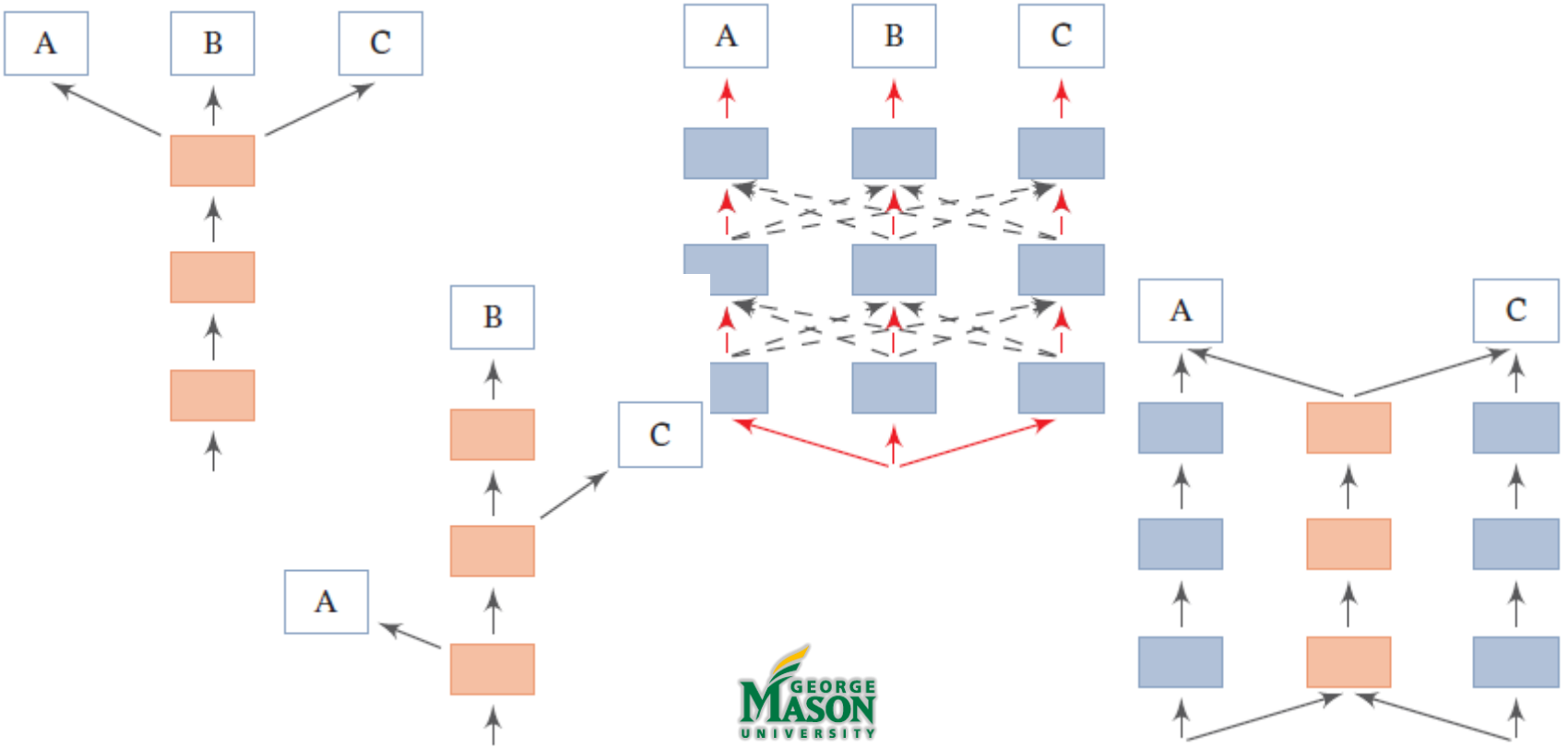
Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data



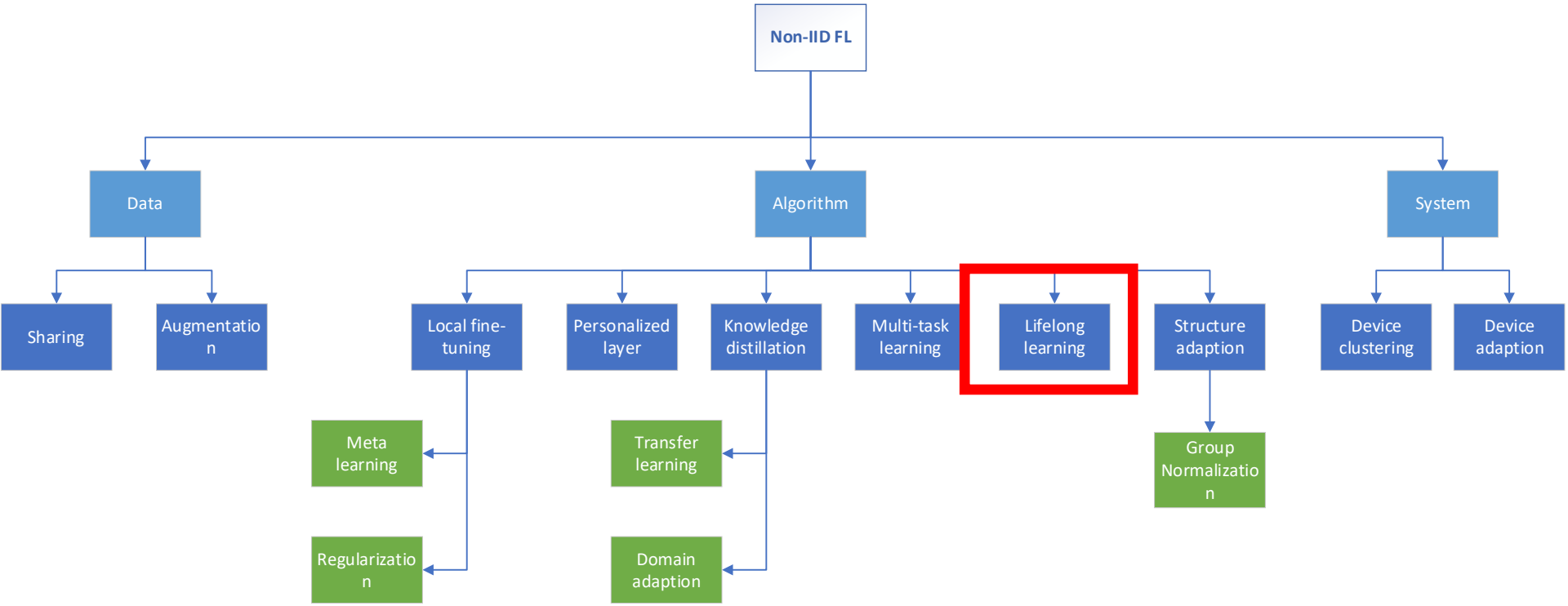
Main Approaches to Handling Non-IID Data

- Multi-task learning
  - [12] Caruana, R., 1997. Multitask learning. Machine learning 28, 41–75.



Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data

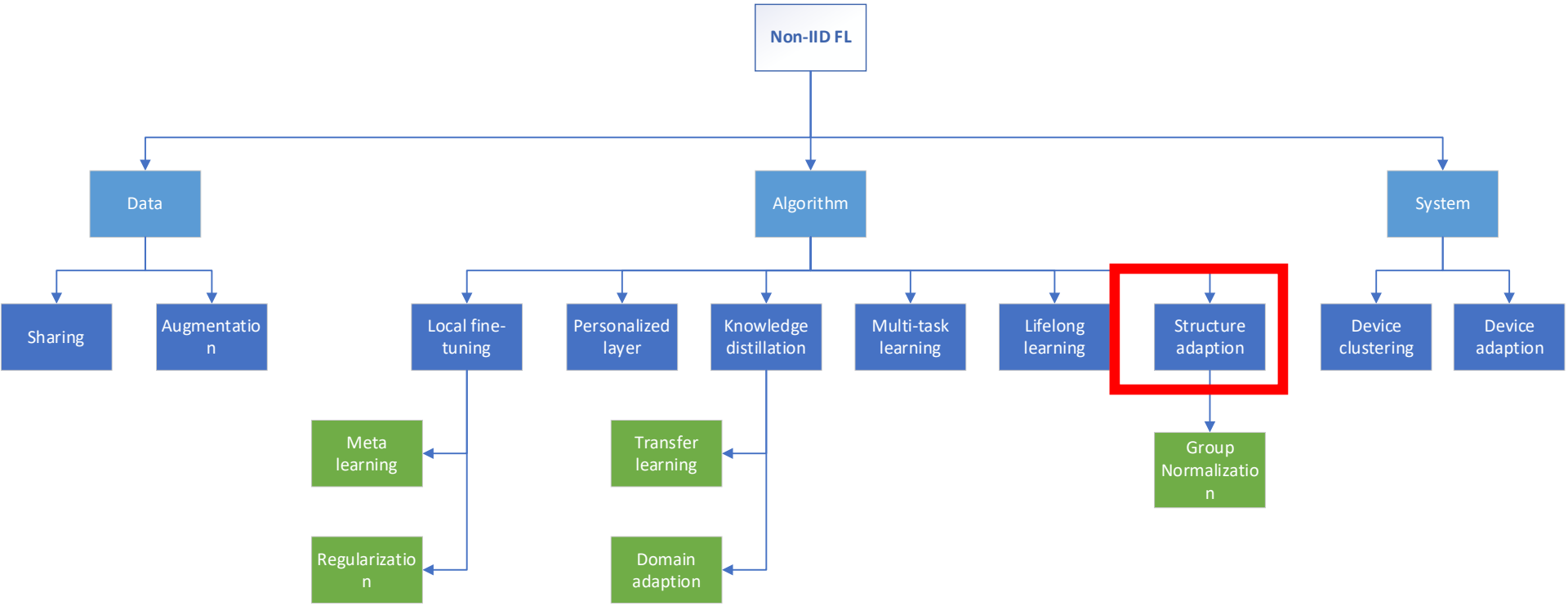


Main Approaches to Handling Non-IID Data

- Lifelong Learning
  - Maintain the model accuracy without forgetting previously learnt tasks
  - [13] Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al., 2017. Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences 114, 3521–3526.
    - The most important parameters for a specific task A are identified.
    - When the model is trained on another task B, the learner will be penalized for changing these parameters

Main Approaches to Handling Non-IID Data

- A summary of existing approaches to addressing Non-IID data

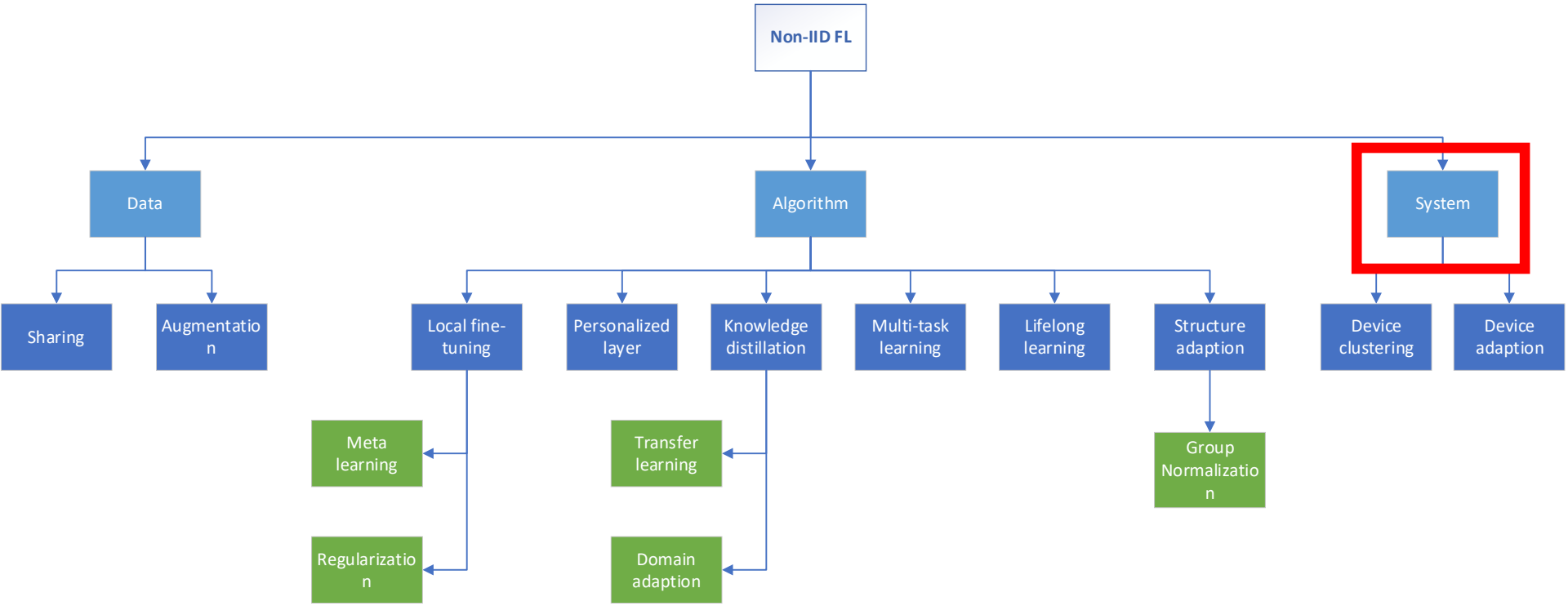


Main Approaches to Handling Non-IID Data

- Structure adaption
  - Group normalization
  - [14] Zhang, Z., Yang, Y., Yao, Z., Yan, Y., Gonzalez, J.E., Mahoney, M.W., 2020b. Improving semi-supervised federated learning by reducing the gradient diversity of models. arXiv preprint arXiv:2008.11364 .
    - Add some group normalization layers into model.

Main Approaches to Handling Non-IID Data

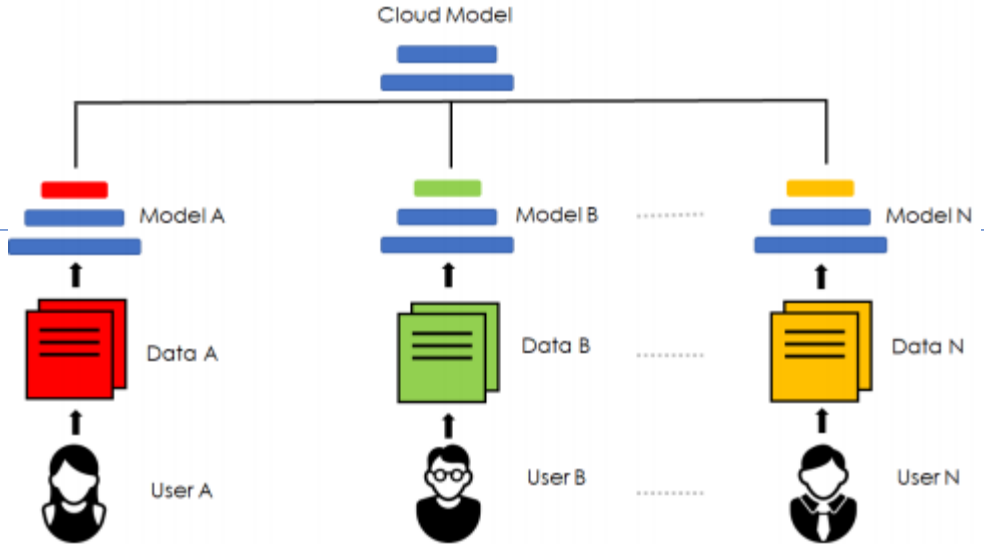
- A summary of existing approaches to addressing Non-IID data



Outline

- Problem Setup
- Main Approaches to Handling Non-IID Data
  - Data based approach
  - Algorithm based approach
  - System based approach
- **Discussion**

Discussion



- Personalization
  - Base layers are shared layers[8]
  - Deep layers are shared layers[9]

- Which layers should be personalized
  - [15] Li X C, Gan L, Zhan D C, et al. Aggregate or Not? Exploring Where to Privatize in DNN Based Federated Learning Under Different Non-IID Scenes[J]. arXiv preprint arXiv:2107.11954, 2021.