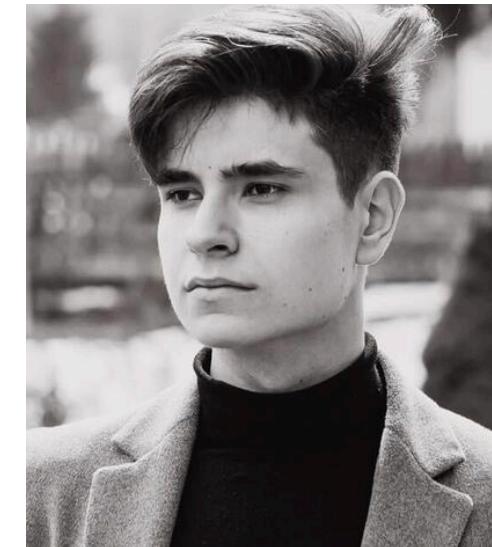


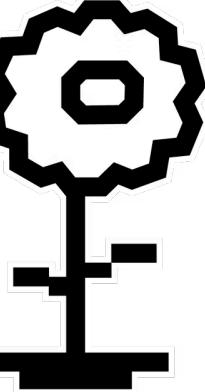
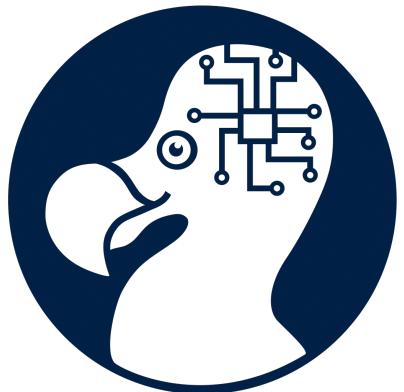
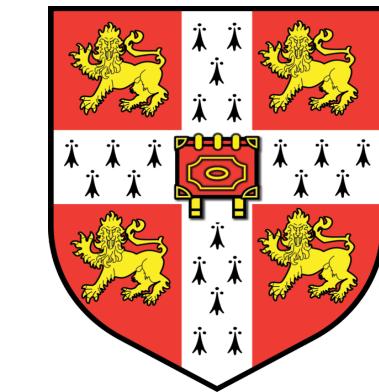
# Can Fair Federated Learning reduce the need for Personalisation?



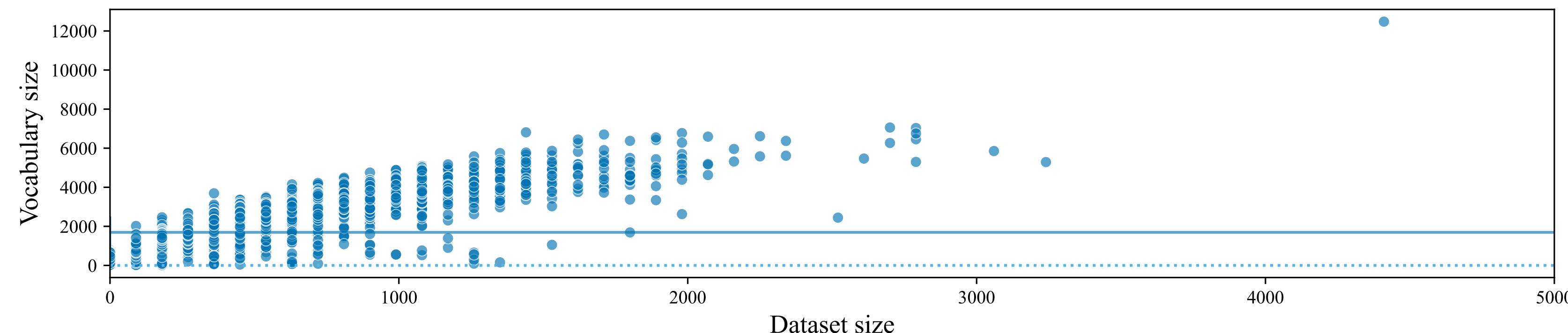
**Alex Jacob, Pedro P. B. Gusmão, Nicholas D. Lane**  
**Cambridge ML Systems**

# Federated Learning

## Heterogeneity

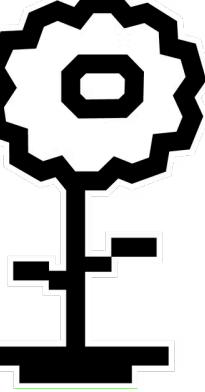
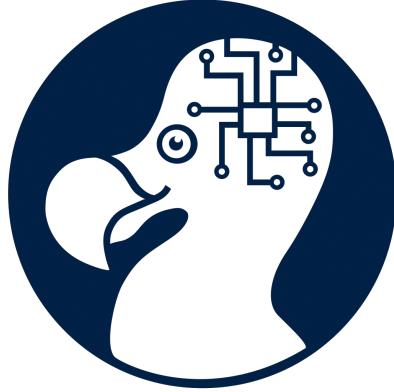
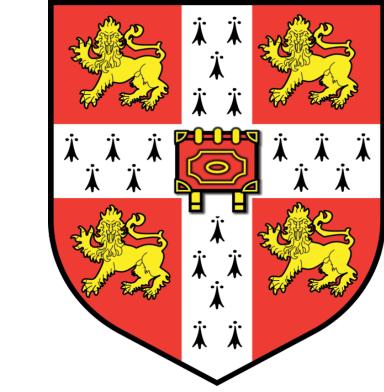


- FL allows edge devices to collaboratively train a **single** model
  - **Without sharing data**
- The shared model does not perform uniformly across clients [Li et al. 2019]
- For some clients it can perform **worse**, than a **fully local** model
  - Such clients lack any **incentive** to participate in the FL process [Salvaging Federated Learning by local adaptation - Yu et al. 2019]



# Addressing Heterogeneity

## Fair FL and Personalisation



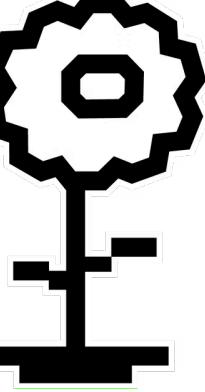
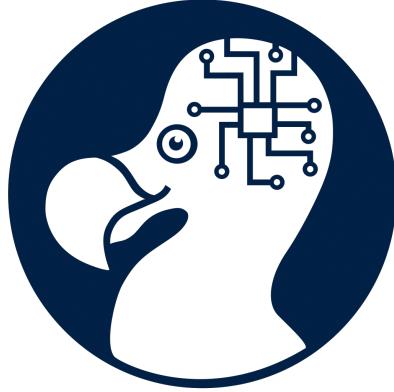
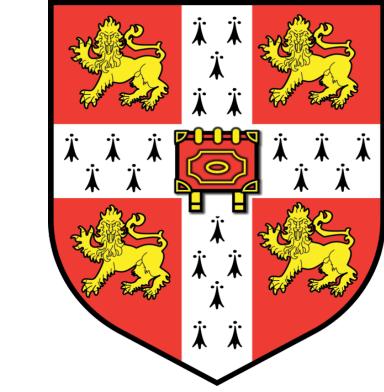
- Fair Federated Learning [Fair resource allocation in federated learning - Li et al. 2019]
  - Attempts to reduce the **accuracy disparity** between clients
  - By focusing on clients with **higher losses** in the objective function

$$\min_w f(w) = \sum_{k=1}^m \frac{p_k}{q+1} F_k^{q+1}(w)$$

- Personalisation
  - Fine-tune the federated model locally while avoiding catastrophic forgetting
  - Using regularisers such as Knowledge Distillation (**KD**) [Hinton et al. 2015], Elastic-weight Consolidation (**EWC**) [Kirkpatrick et al. 2017], or Freebase(**FB**)

# Addressing Heterogeneity

## Fair FL and Personalisation



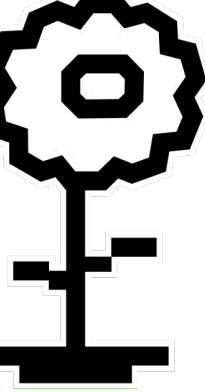
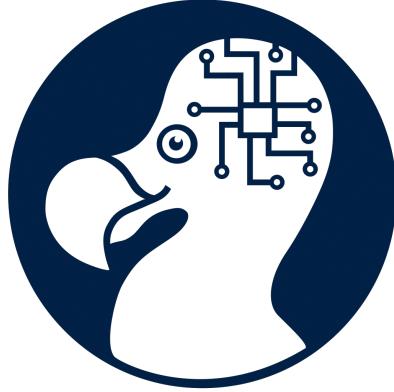
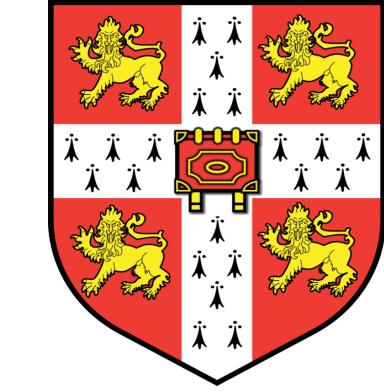
- Fair Federated Learning [Fair resource allocation in federated learning - Li et al. 2019]
  - Attempts to reduce the **accuracy disparity** between clients
  - By focusing on clients with **higher losses** in the objective function

$$\min_w f(w) = \sum_{k=1}^m \frac{p_k}{q+1} F_k^{q+1}(w)$$

- Personalisation
  - Fine-tune the federated model locally while avoiding catastrophic forgetting
  - Using regularisers such as Knowledge Distillation (**KD**) [Hinton et al. 2015], Elastic-weight Consolidation (**EWC**) [Kirkpatrick et al. 2017], or Freebase(**FB**)

# Addressing Heterogeneity

## Fair FL and Personalisation



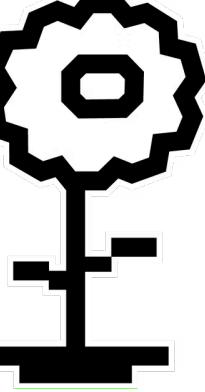
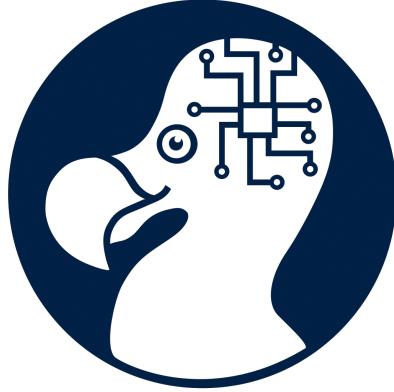
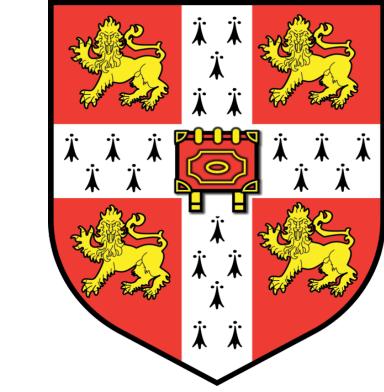
- Fair Federated Learning [Fair resource allocation in federated learning - Li et al. 2019]
  - Attempts to reduce the **accuracy disparity** between clients
  - By focusing on clients with **higher losses** in the objective function

$$\min_w f(w) = \sum_{k=1}^m \frac{p_k}{q+1} F_k^{q+1}(w)$$

- Personalisation
  - Fine-tune the federated model locally while avoiding catastrophic forgetting
  - Using regularisers such as Knowledge Distillation (**KD**) [Hinton et al. 2015], Elastic-weight Consolidation (**EWC**) [Kirkpatrick et al. 2017], or Freebase(**FB**)

# Addressing Heterogeneity

## Fair FL and Personalisation



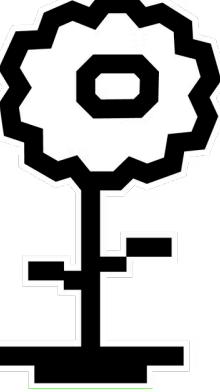
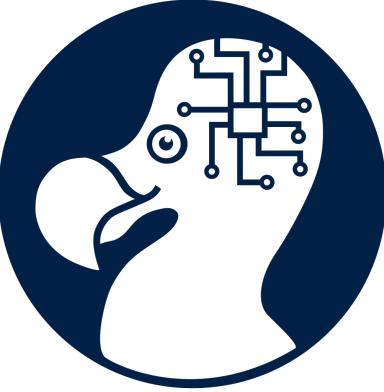
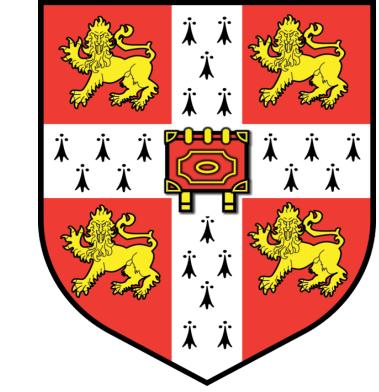
- Fair Federated Learning [Fair resource allocation in federated learning - Li et al. 2019]
  - Attempts to reduce the **accuracy disparity** between clients
  - By focusing on clients with **higher losses** in the objective function

$$\min_w f(w) = \sum_{k=1}^m \frac{p_k}{q+1} F_k^{q+1}(w)$$

- Personalisation
  - Fine-tune the federated model locally while avoiding catastrophic forgetting
  - Using regularisers such as Knowledge Distillation (**KD**) [Hinton et al. 2015], Elastic-weight Consolidation (**EWC**) [Kirkpatrick et al. 2017], or Freebase(**FB**)

# Addressing Heterogeneity

## Fair FL and Personalisation



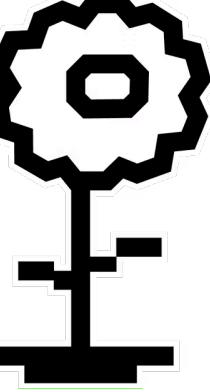
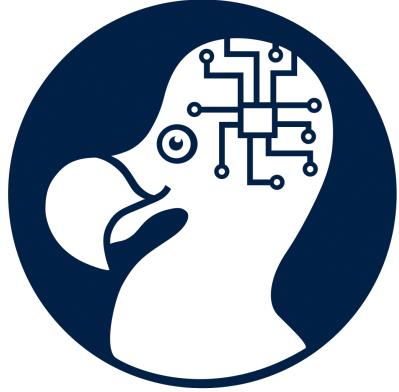
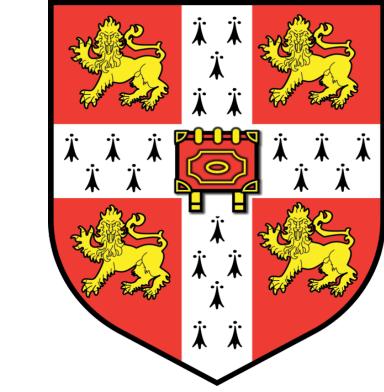
- Fair Federated Learning [Fair resource allocation in federated learning - Li et al. 2019]
  - Attempts to reduce the **accuracy disparity** between clients
  - By focusing on clients with **higher losses** in the objective function

$$\min_w f(w) = \sum_{k=1}^m \frac{p_k}{q+1} F_k^{q+1}(w)$$

- Personalisation
  - Fine-tune the federated model locally while avoiding catastrophic forgetting
  - Using regularisers such as Knowledge Distillation (**KD**) [Hinton et al. 2015], Elastic-weight Consolidation (**EWC**) [Kirkpatrick et al. 2017], or Freebase(**FB**)

# Motivation

## Reducing the need for personalisation



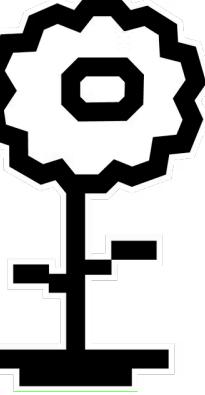
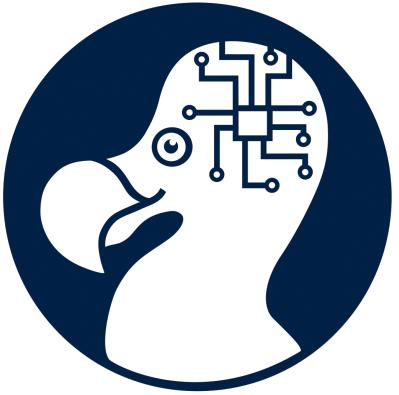
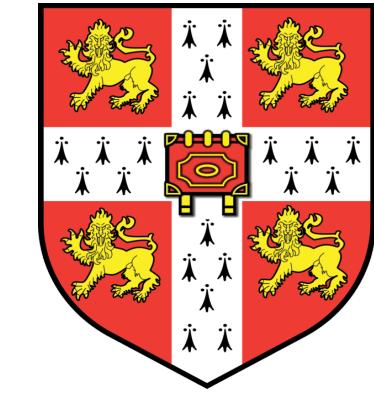
- Personalisation is costly and requires the final clients to be able to train the model rather than merely running inference
- Requires maintaining state on potentially millions of devices

# Our Contribution

- We show Fair FL (FFL) does **not** reduce the need for **nor** benefit later personalisation
  - By looking at relative accuracy: the difference in accuracy between a federated and fully local model
- We propose using regularisers that anticipate personalisation in **Personalisation-aware FL**
- We show Personalisation-aware FL (PaFL) **outperforms** FFL in our simulations

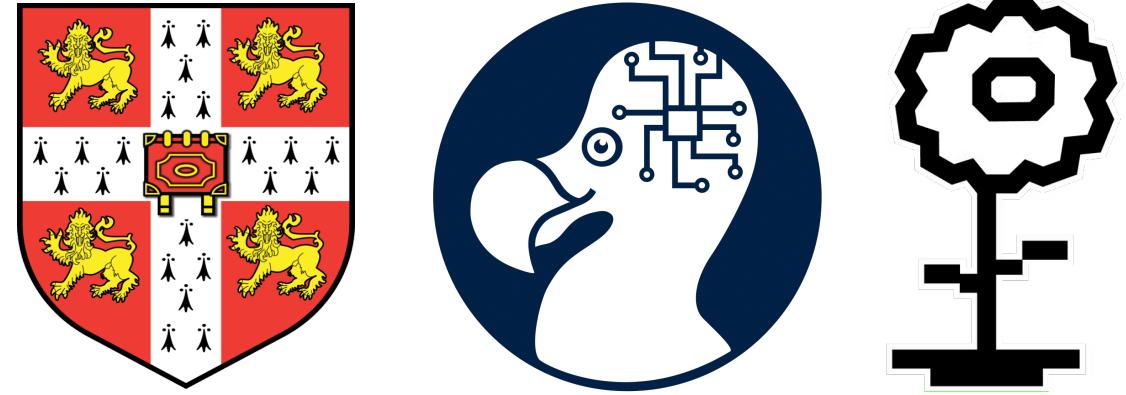
# Personalisation-aware FL

PaFL

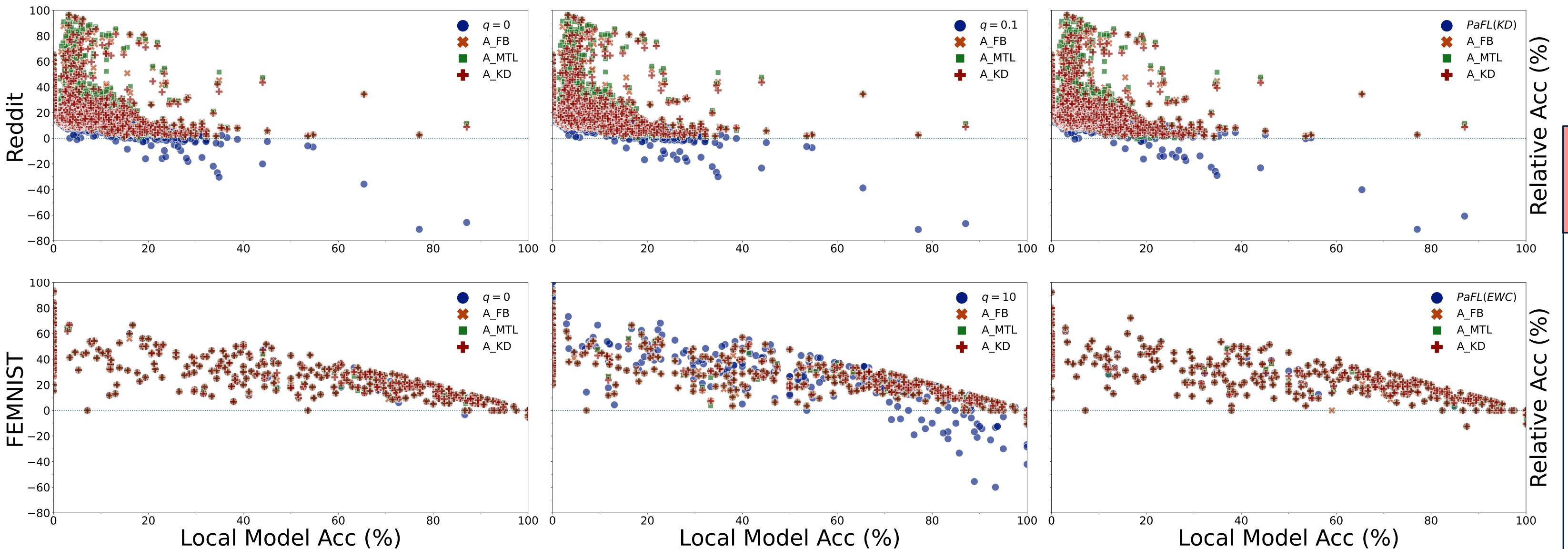


- Fair FL flattens the accuracy distribution and may harm the best performing clients
- **PaFL** as an alternative
  - Inspired by Quantisation-aware training
  - Use regularisers like KD, EWC, FB **during** FL training
  - Allows their weight to **vary across rounds**
- Maintains performance of the FL model on avg while learning from heterogeneous clients
- May have synergistic effects with the final personalisation regulariser

# Results

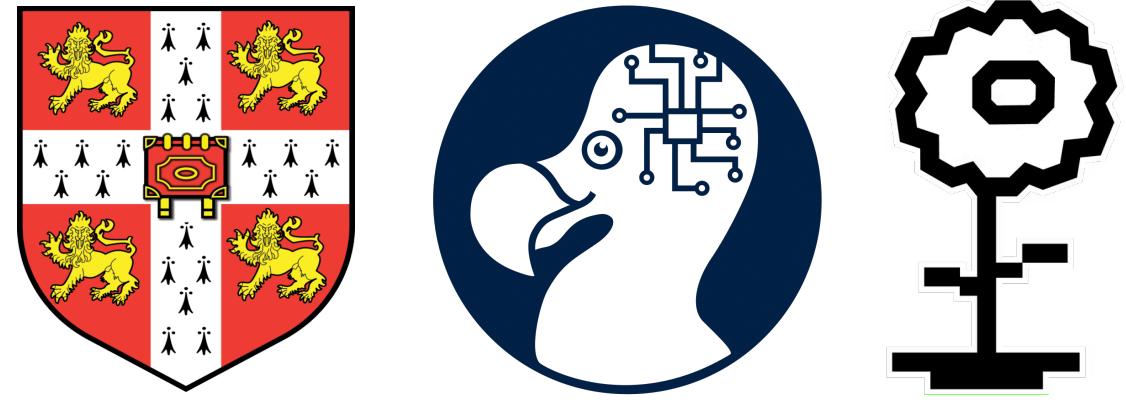


- **Fair FL:**
  - Slightly **decreases** avg relative accuracy on Reddit
  - **Doubles** underperformers on FEMNIST with KD
    - Despite **improving** avg relative accuracy

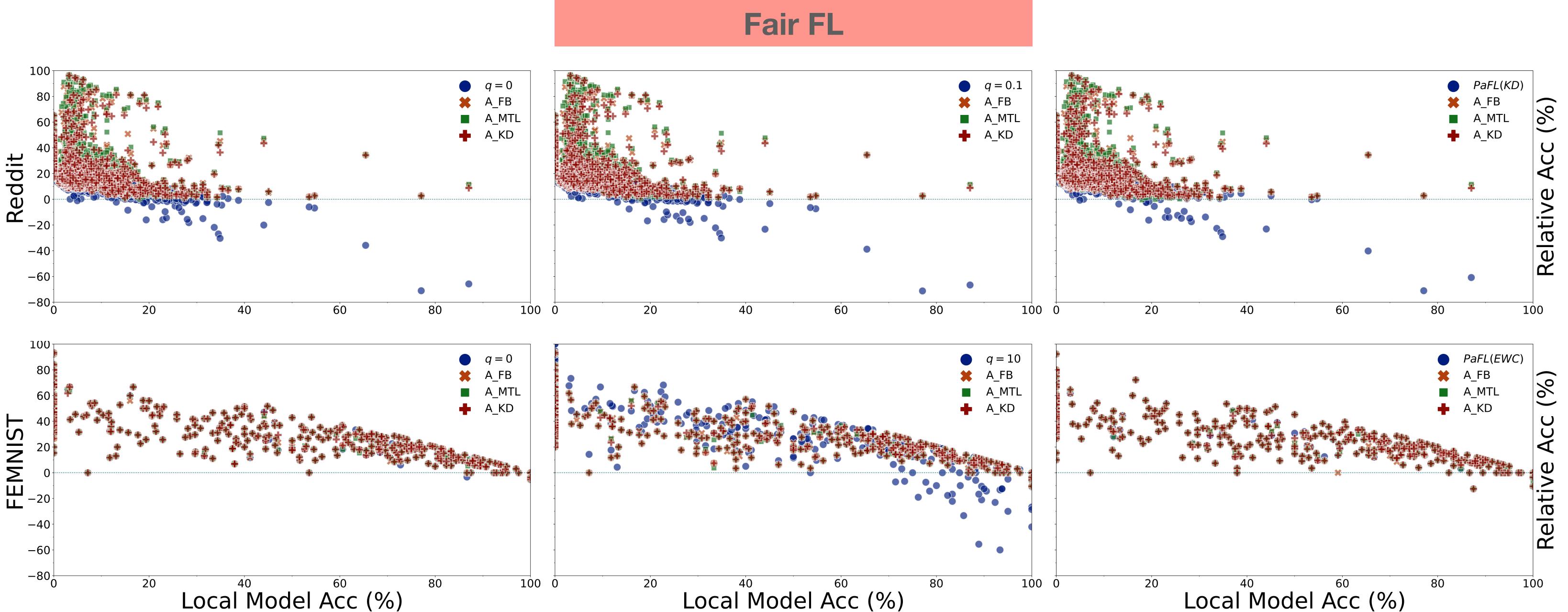


- **PaFL:**
  - **Halves** underperforming clients on Reddit
  - Achieves **similar** average relative accuracy on FEMNIST when using EWC
    - With a slight reduction in underperforming clients which **persists** after personalisation

# Results

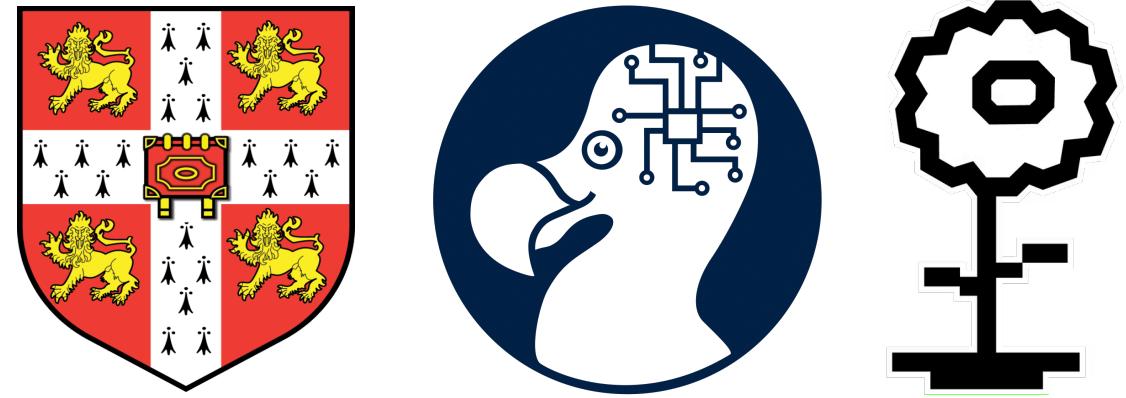


- **Fair FL:**
  - Slightly **decreases** avg relative accuracy on Reddit
  - **Doubles** underperformers on FEMNIST with KD
    - Despite **improving** avg relative accuracy

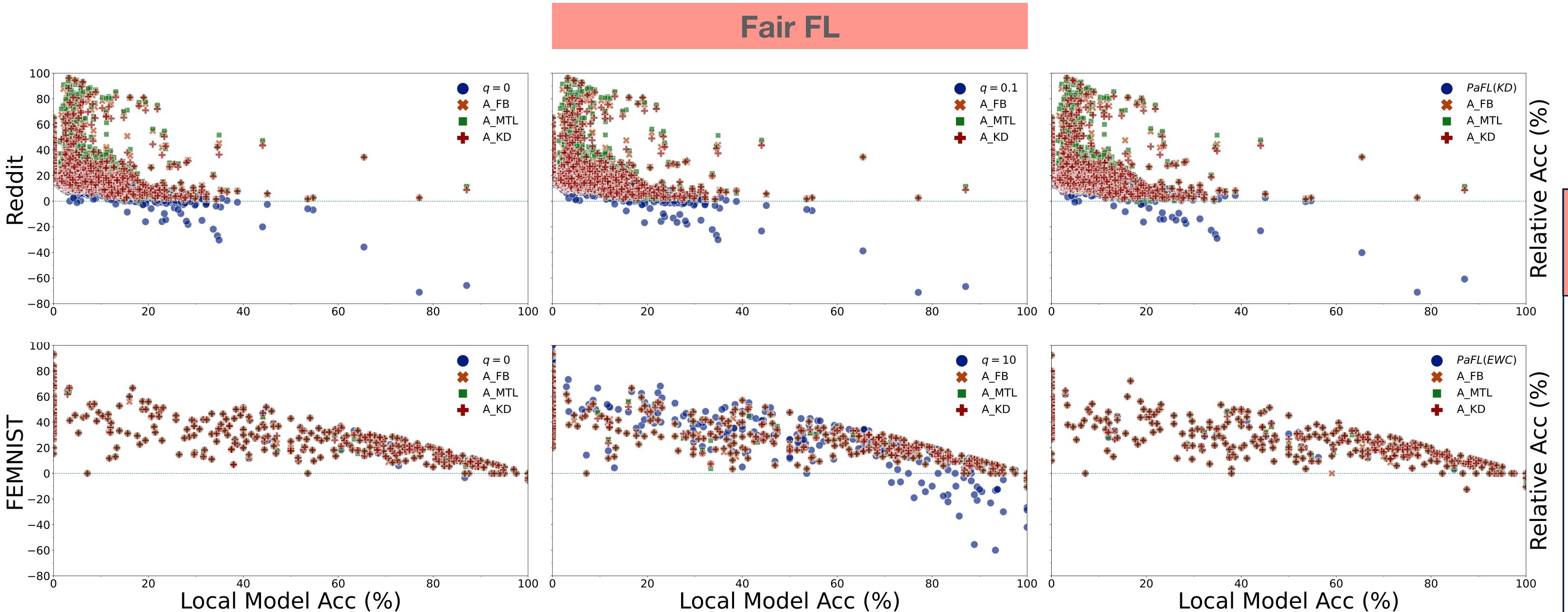


- **PaFL:**
  - **Halves** underperforming clients on Reddit
  - Achieves **similar** average relative accuracy on FEMNIST when using EWC
    - With a slight reduction in underperforming clients which **persists** after personalisation

# Results

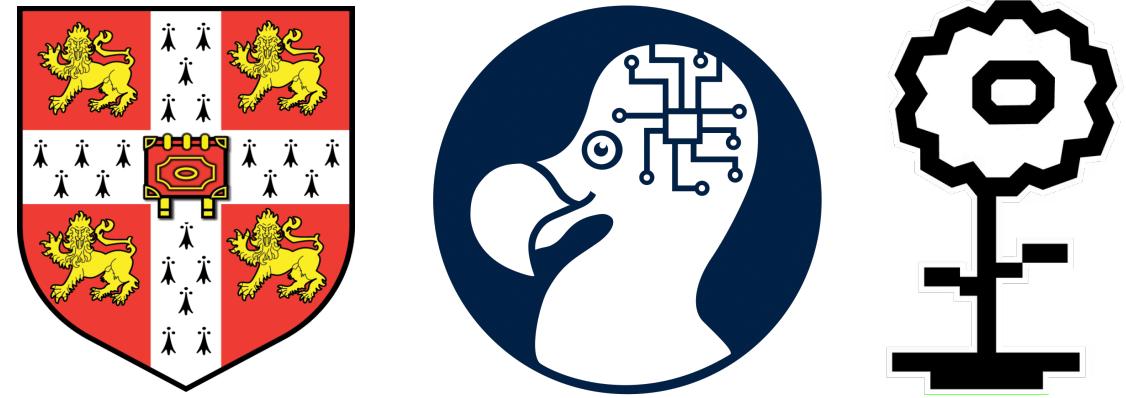


- **Fair FL:**
  - Slightly **decreases** avg relative accuracy on Reddit
  - **Doubles** underperformers on FEMNIST with KD
    - Despite **improving** avg relative accuracy

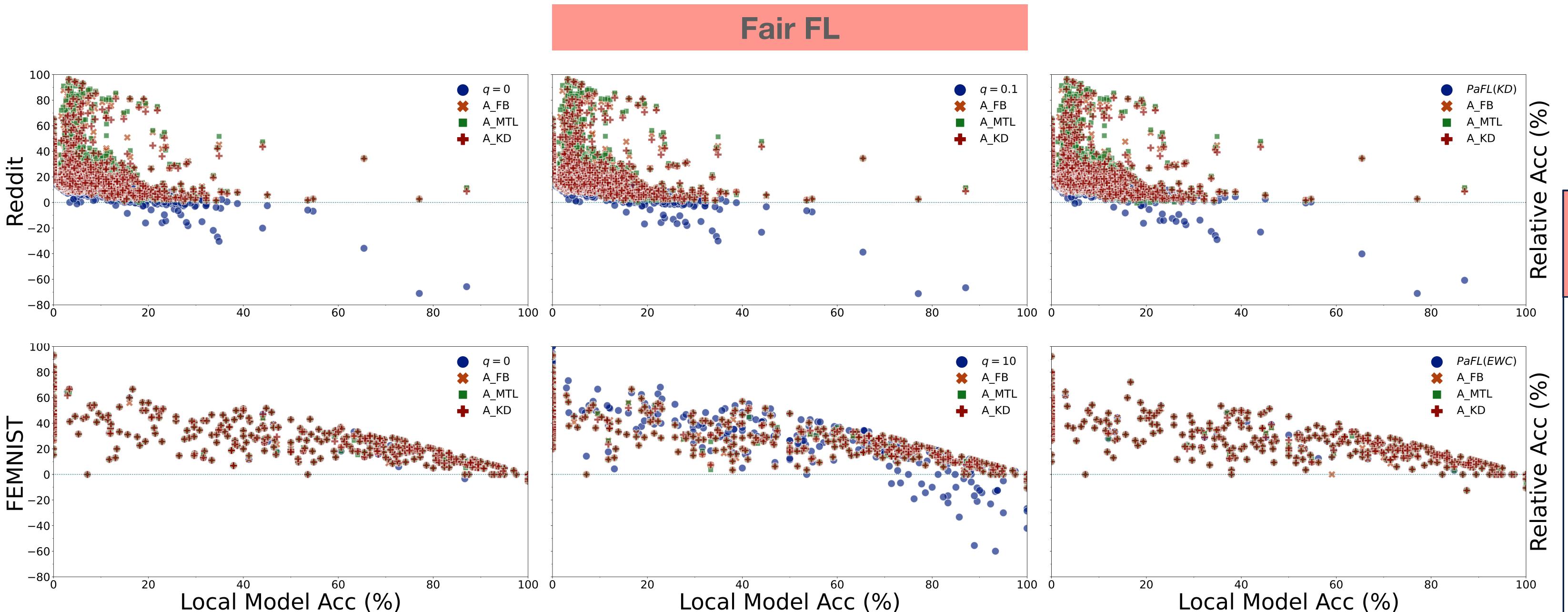


- **PaFL:**
  - **Halves** underperforming clients on Reddit
  - Achieves **similar** average relative accuracy on FEMNIST when using EWC
    - With a slight reduction in underperforming clients which **persists** after personalisation

# Results

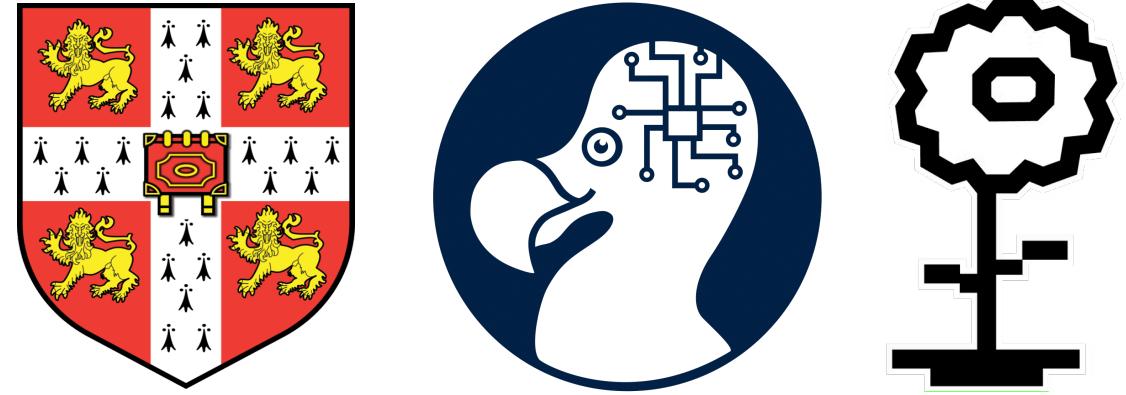


- **Fair FL:**
  - Slightly **decreases** avg relative accuracy on Reddit
  - **Doubles** underperformers on FEMNIST with KD
    - Despite **improving** avg relative accuracy



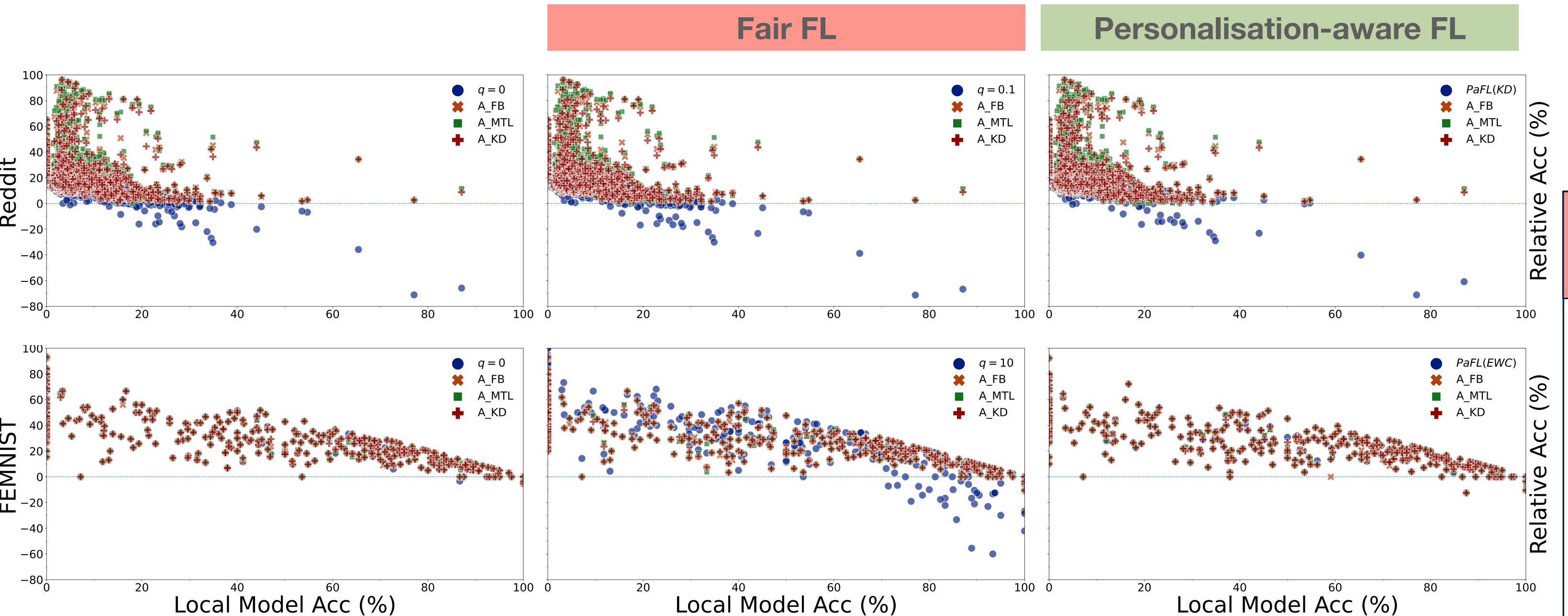
- **PaFL:**
  - **Halves** underperforming clients on Reddit
  - Achieves **similar** average relative accuracy on FEMNIST when using EWC
    - With a slight reduction in underperforming clients which **persists** after personalisation

# Results

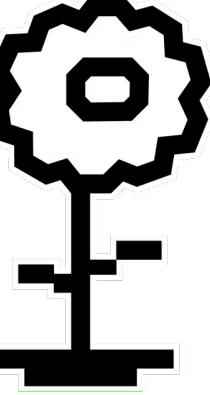
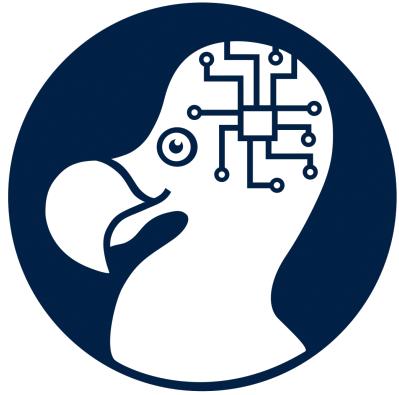
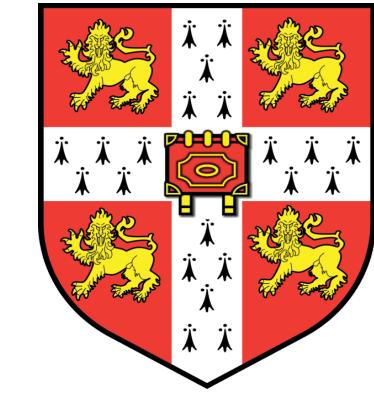


- **Fair FL:**
  - Slightly **decreases** avg relative accuracy on Reddit
  - **Doubles** underperformers on FEMNIST with KD
    - Despite **improving** avg relative accuracy

- **PaFL:**
  - **Halves** underperforming clients on Reddit
  - Achieves **similar** average relative accuracy on FEMNIST when using EWC
    - With a slight reduction in underperforming clients which **persists** after personalisation

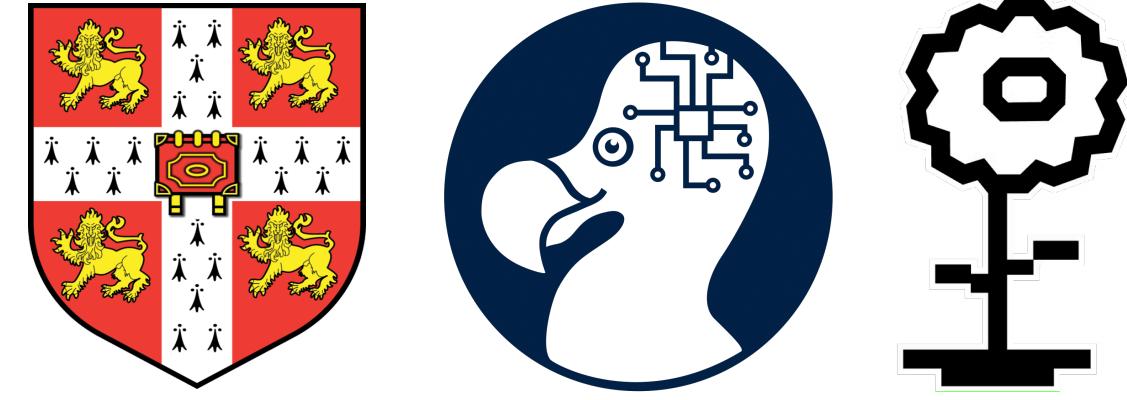


# Conclusion



- **Fair FL** is capable of reducing the variance of the accuracy distribution over clients
  - This leads to decreased performance on clients capable of training a very high quality local model
  - Thus, it **hurts** the **relative accuracy** distribution by skewing it towards negative values
  - It shows **no benefits** to later personalisation
- **Personalisation-aware Federate Learning (PaFL)** provides an alternative which:
  - Trains on heterogeneous clients while maintaining performance on the federated distribution
  - Tends to lead to equivalent or **higher** accuracy on average
  - May **reduce** underperforming clients by up to **50%** and potentially enhance later personalisation

# References



- Li, T., Sanjabi, M., Beirami, A., & Smith, V. (2019). Fair resource allocation in federated learning. *arXiv preprint arXiv:1905.10497*.
- Yu, T., Bagdasaryan, E., & Shmatikov, V. (2020). Salvaging federated learning by local adaptation. *arXiv preprint arXiv:2002.04758*.
- Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.
- Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13), 3521-3526.