

# A Mutual Information Perspective on Federated Contrastive Learning

Christos Louizos ♡, Matthias Reisser, Denis Korzhenkov

♡ Eng., Senior Staff, Qualcomm Technologies Netherlands B.V.

{clouizos,mreisser,dkorzhenkov}@qti.qualcomm.com

† Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Snapdragon and Qualcomm branded products are products of Qualcomm Technologies, Inc. and/or its subsidiaries.

# Self-supervised learning in the federated context

# Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?

# Self-supervised learning in the federated context

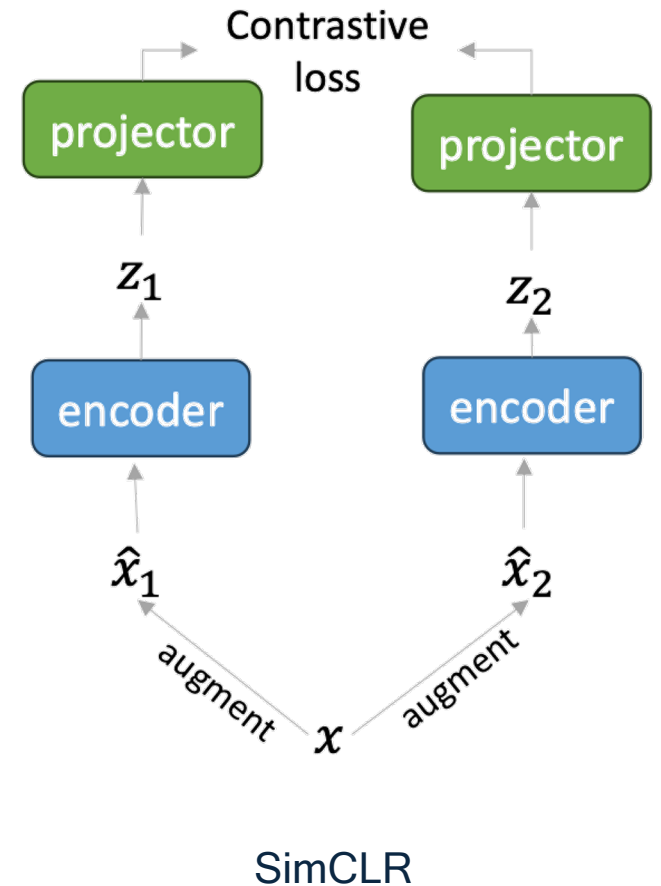
- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset

# Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand

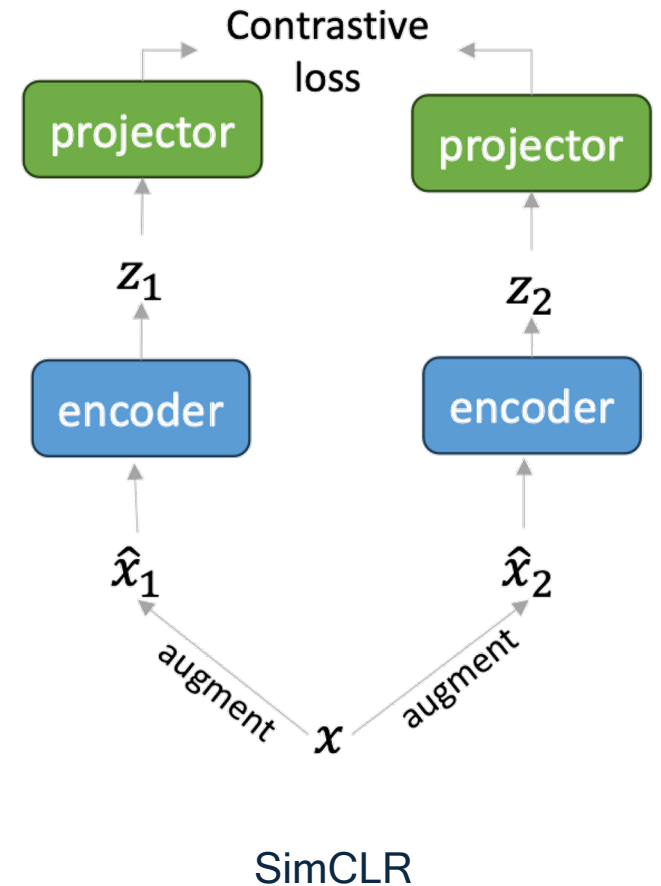
# Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand
- SimCLR is a popular centralized self-supervised paradigm
  - How can we extend it from first principles to the federated setting?



# Self-supervised learning in the federated context

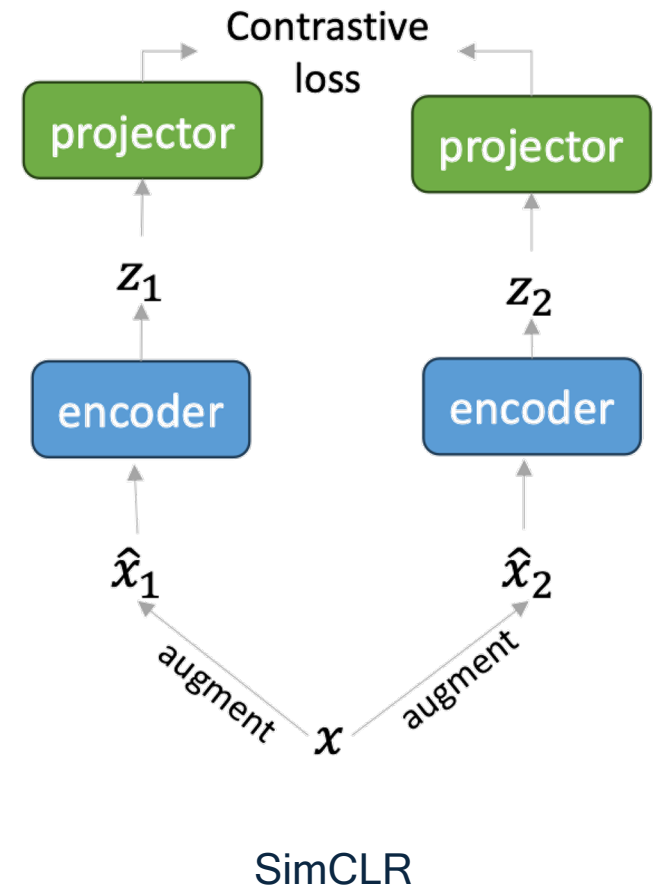
- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand
- SimCLR is a popular centralized self-supervised paradigm
  - How can we extend it from first principles to the federated setting?
- This work:  
We extend SimCLR to FL through a mutual-information perspective



# Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand
- SimCLR is a popular centralized self-supervised paradigm
  - How can we extend it from first principles to the federated setting?
- This work:  
We extend SimCLR to FL through a mutual-information perspective

*Bonus: extension also covers the semi-supervised setting*



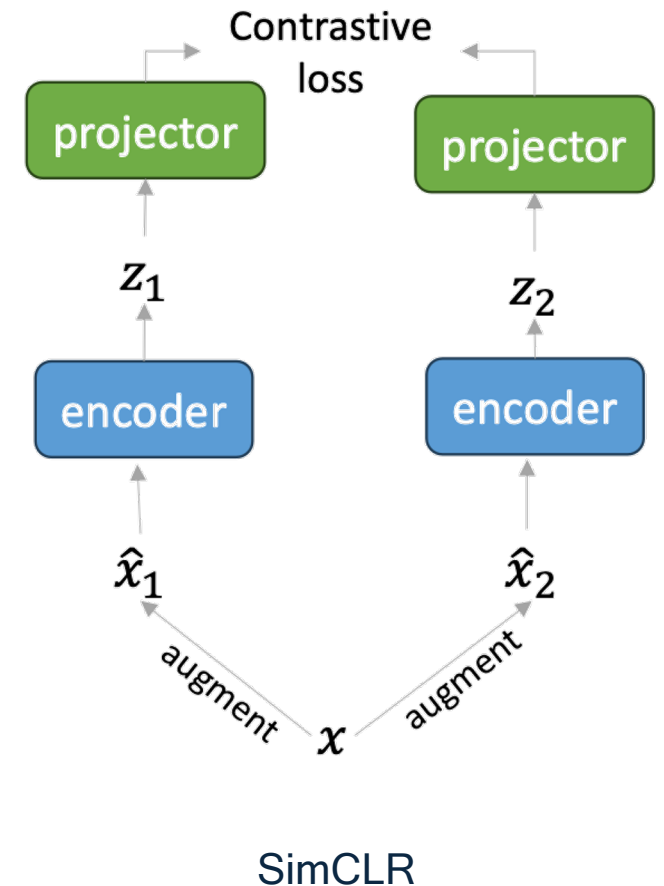


# Self-supervised learning in the federated context

- Obtaining labelled data in federated learning (FL) is tricky
  - Who provides the labels at the client level?
- Commonly, we observe
  - Client have unlabeled data
  - Server has a (small) labelled / curated dataset
- Solution: self-supervised federated learning
  - Use the federation for self-supervised training
  - Fine-tune using a central labelled dataset for the task at hand
- SimCLR is a popular centralized self-supervised paradigm
  - How can we extend it from first principles to the federated setting?
- This work:  
We extend SimCLR to FL through a mutual-information perspective

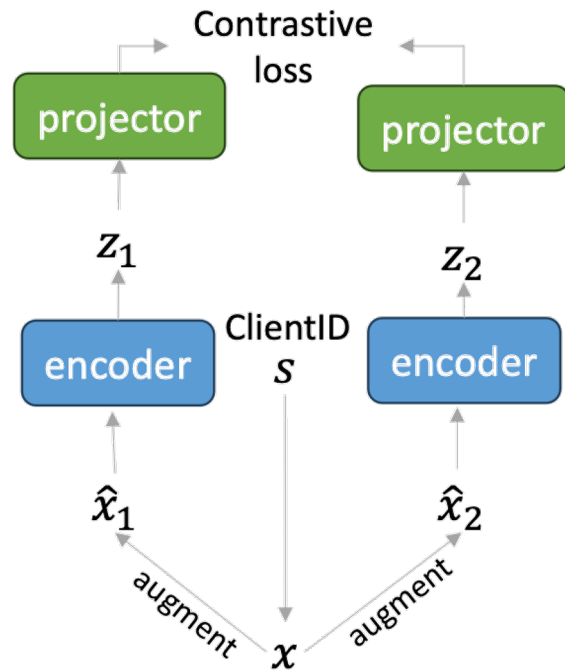
*Bonus:* extension also covers the semi-supervised setting

*Second Bonus:* insights translate to other self-supervised methods



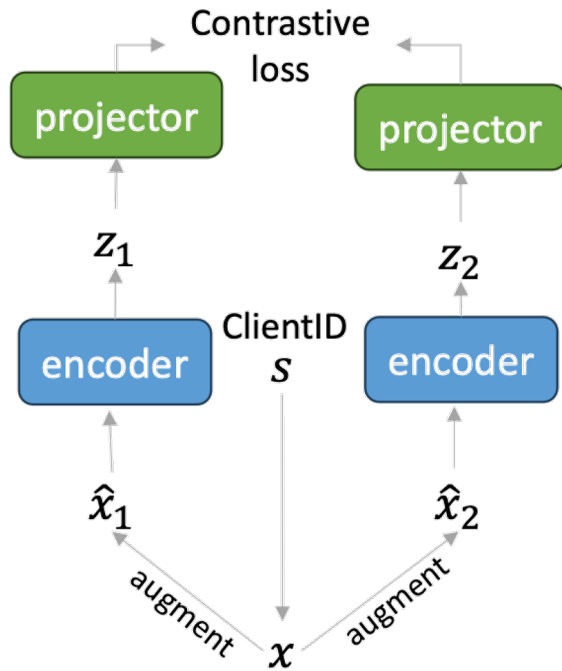
# The SimCLR variants considered

# The SimCLR variants considered

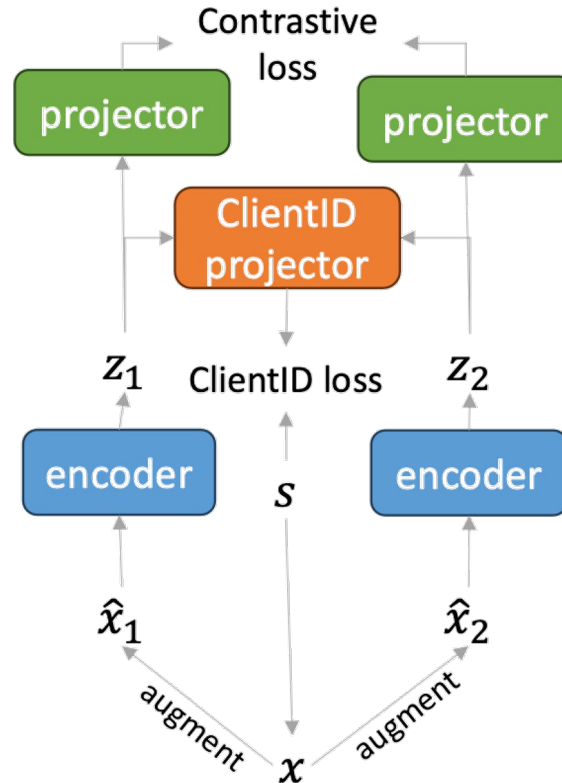


**Local SimCLR:** each client applies SimCLR on their own data, thus the federation implicitly optimizes a lower bound to the local  $I(z_1; z_2 | s)$

# The SimCLR variants considered

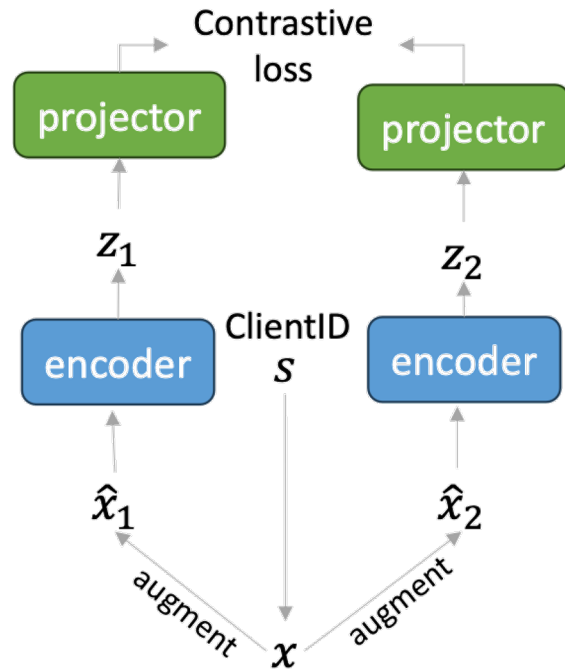


**Local SimCLR:** each client applies SimCLR on their own data, thus the federation implicitly optimizes a lower bound to the local  $I(z_1; z_2 | s)$

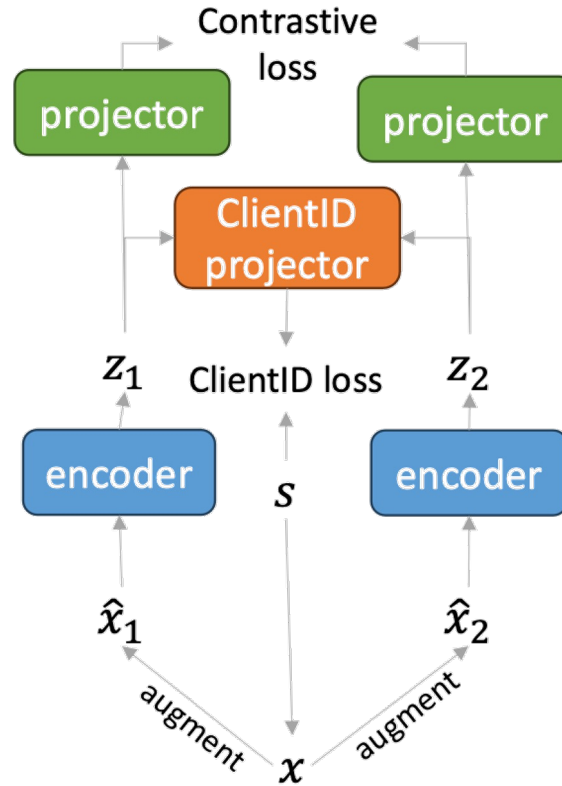


**Federated SimCLR:** each client also optimizes a client classifier, thus the federation implicitly optimizes a lower bound to the global  $I(z_1; z_2)$

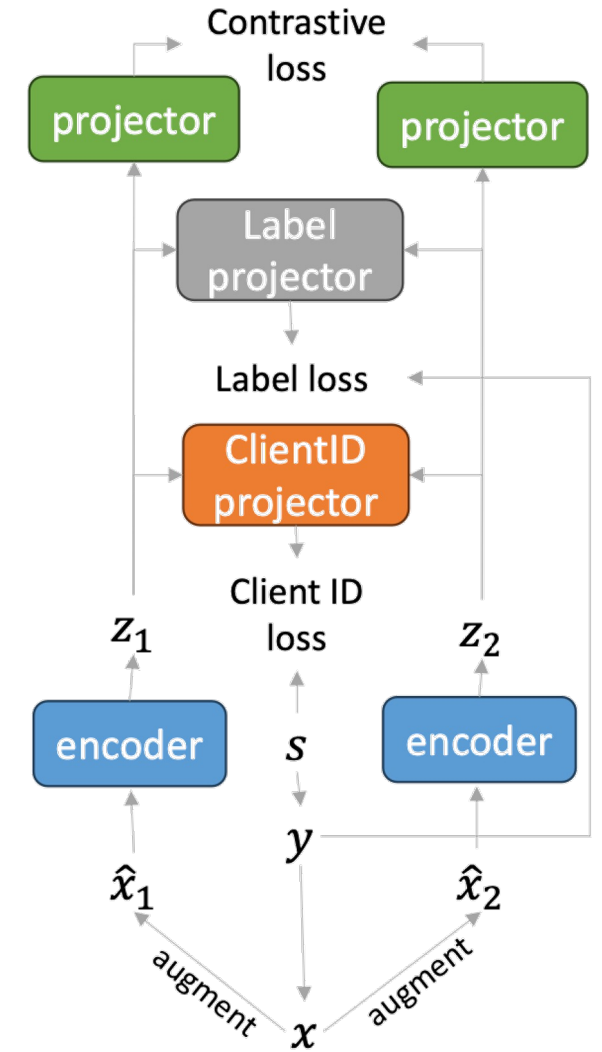
# The SimCLR variants considered



**Local SimCLR:** each client applies SimCLR on their own data, thus the federation implicitly optimizes a lower bound to the local  $I(z_1; z_2|s)$



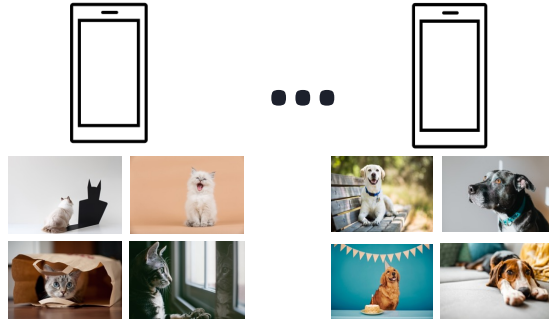
**Federated SimCLR:** each client also optimizes a client classifier, thus the federation implicitly optimizes a lower bound to the global  $I(z_1; z_2)$



**Supervised federated SimCLR:** with an additional label classifier and label dependent contrastive learning, we obtain a label-informed variant that also optimizes a lower bound to the global  $I(z_1; z_2)$

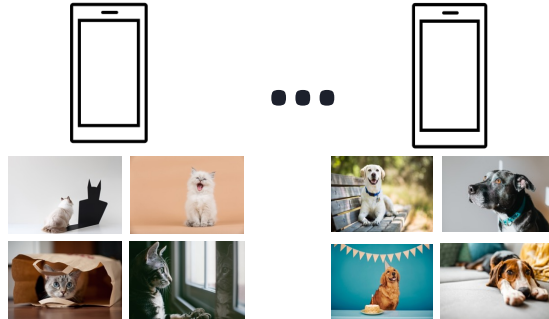
# Federated Learning: the non-i.i.d. generative process

# Federated Learning: the non-i.i.d. generative process



- **Label-skew**: the most common non-iid setting assumed in FL
  - Each client has a different class marginal,  $p(y|s)$
  - The conditional feature distribution,  $p(x|y)$ , is the same for all clients

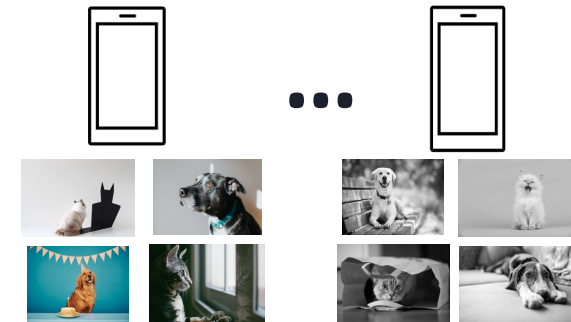
# Federated Learning: the non-i.i.d. generative process



- **Label-skew:** the most common non-iid setting assumed in FL
  - Each client has a different class marginal,  $p(y|s)$
  - The conditional feature distribution,  $p(x|y)$ , is the same for all clients

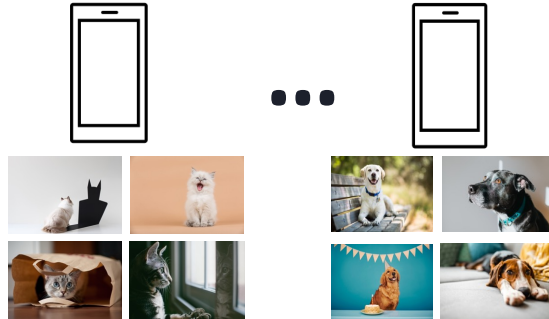
- **Covariate shift:** feature noise, independent of the label

- Each client has the same local distribution over classes,  $p(y)$
- The conditional distribution over features,  $p(x|y, s)$ , varies due to, e.g., sensor noise





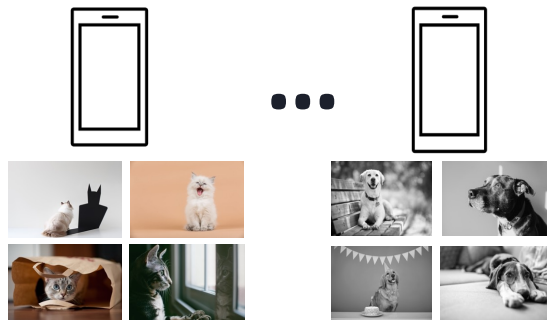
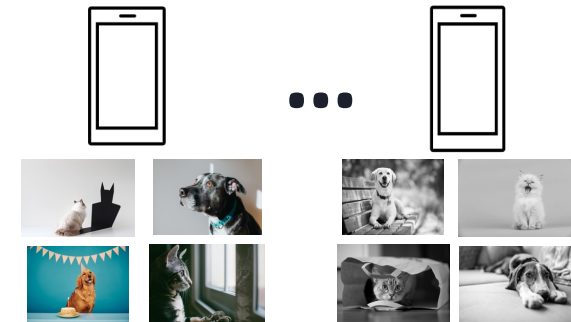
# Federated Learning: the non-i.i.d. generative process



- **Label-skew:** the most common non-iid setting assumed in FL
  - Each client has a different class marginal,  $p(y|s)$
  - The conditional feature distribution,  $p(x|y)$ , is the same for all clients

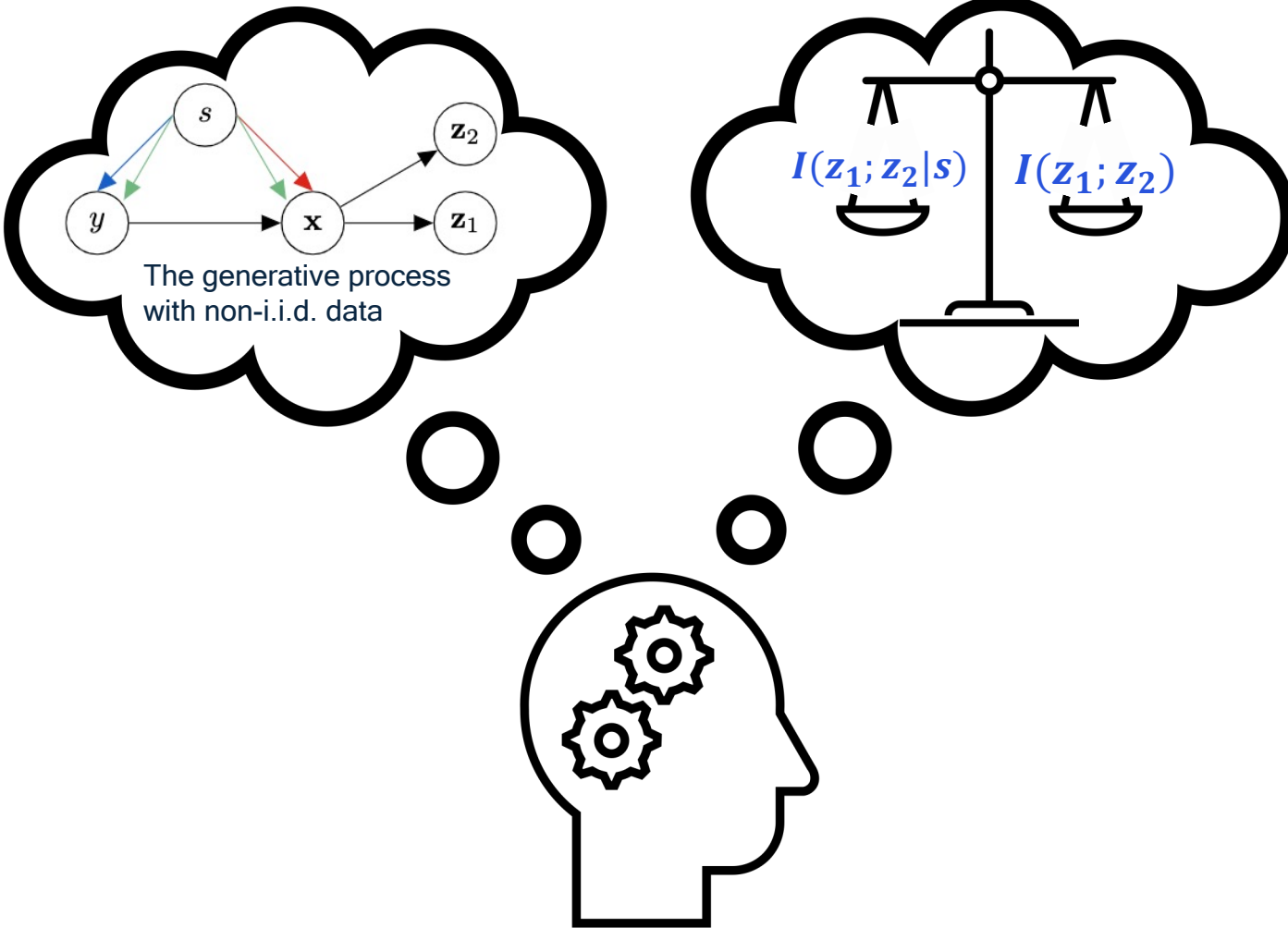
- **Covariate shift:** feature noise, independent of the label

- Each client has the same local distribution over classes,  $p(y)$
- The conditional distribution over features,  $p(x|y, s)$ , varies due to, e.g., sensor noise



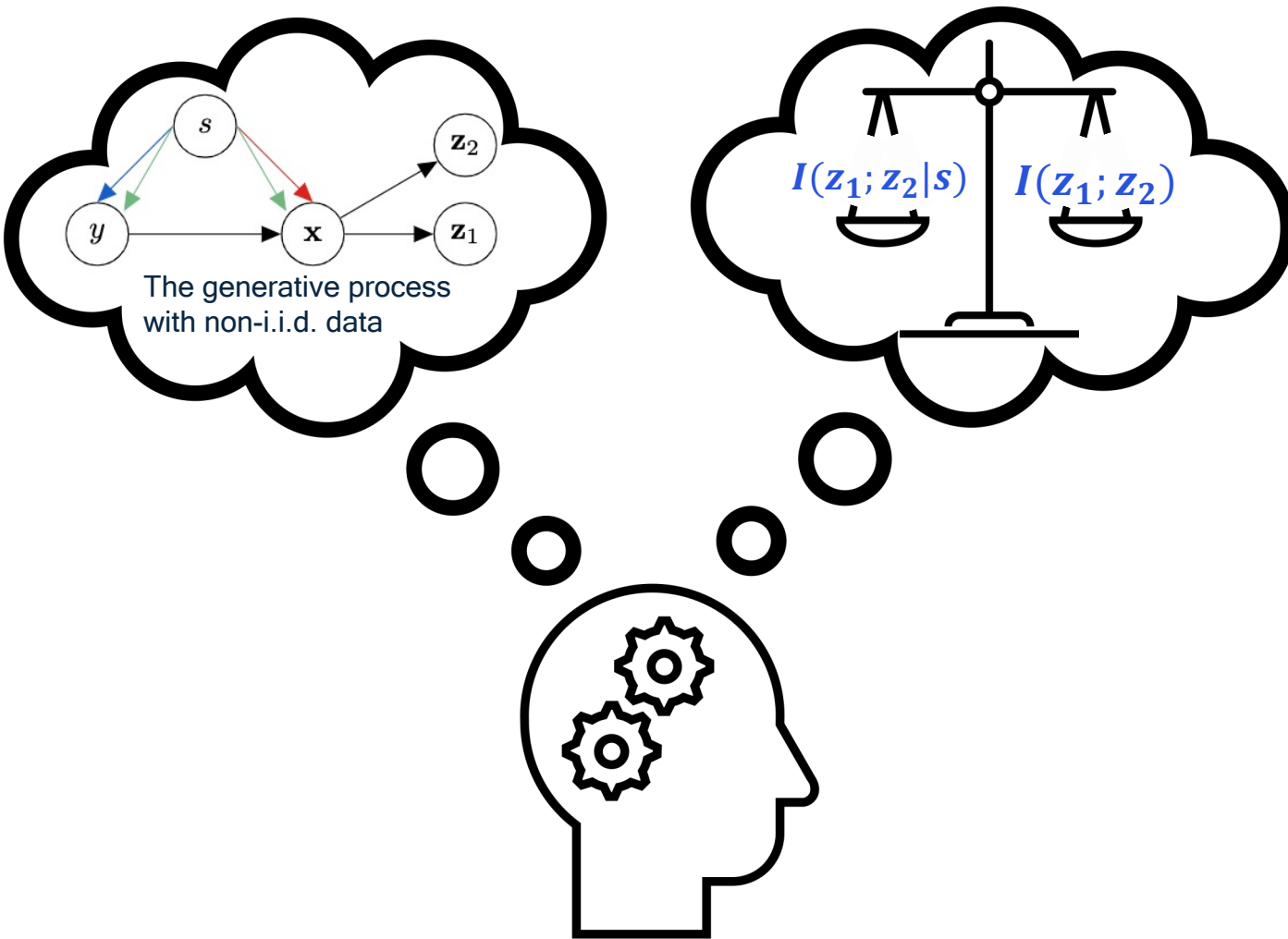
- **Joint shift:** a mixture of the two
  - Both the label marginal and feature distribution vary per client,  $p(y|s)p(x|y, s)$

# The effects of non-i.i.d.-ness on SimCLR



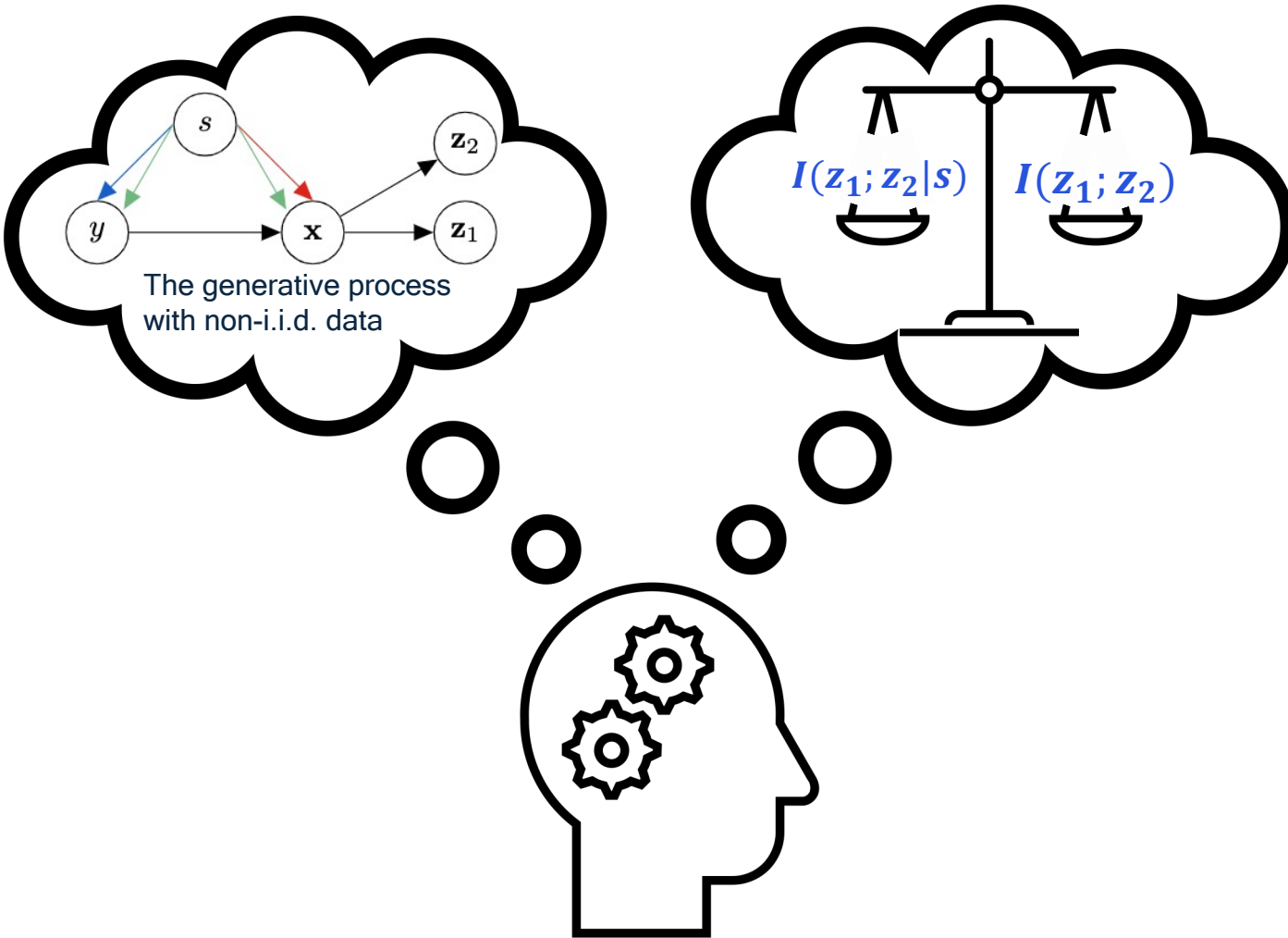
- What are the effects of data non-i.i.d.-ness?
  - Use local  $I(z_1; z_2 | s)$  or global  $I(z_1; z_2)$ ?
  - The client classification task is what separates them

# The effects of non-i.i.d.-ness on SimCLR



- What are the effects of data non-i.i.d.-ness?
  - Use local  $I(z_1; z_2 | s)$  or global  $I(z_1; z_2)$ ?
  - The client classification task is what separates them
- When we have label skew, the client classification task is beneficial
  - We prove that it maximizes a lower bound to the mutual information between the representations and the unknown ground truth label

# The effects of non-i.i.d.-ness on SimCLR



- What are the effects of data non-i.i.d.-ness?
  - Use local  $I(z_1; z_2 | s)$  or global  $I(z_1; z_2)$ ?
  - The client classification task is what separates them
- When we have label skew, the client classification task is beneficial
  - We prove that it maximizes a lower bound to the mutual information between the representations and the unknown ground truth label
- When we have covariate shift, it can be detrimental
  - It encourages storing in the representations irrelevant, for the downstream task, information

# Experimental results

# Experimental results

## Results in the unsupervised case

Method	CIFAR 10			CIFAR 100		
	Label skew	Covariate shift	Joint shift	Label skew	Covariate shift	Joint shift
Local SimCLR	79.4 $\pm$ 0.2	<b>74.3</b> $\pm$ 0.3	71.0 $\pm$ 0.4	42.2 $\pm$ 0.2	<b>41.2</b> $\pm$ 0.2	38.1 $\pm$ 0.3
Federated SimCLR	<b>85.0</b> $\pm$ 0.2	73.8 $\pm$ 0.2	<b>74.8</b> $\pm$ 0.5	<b>48.5</b> $\pm$ 0.1	39.5 $\pm$ 0.2	<b>43.1</b> $\pm$ 0.2
Spectral CL	76.5 $\pm$ 1.1	<b>73.5</b> $\pm$ 0.4	68.2 $\pm$ 0.6	33.3 $\pm$ 6.0	<b>33.6</b> $\pm$ 2.3	<b>29.6</b> $\pm$ 6.2
Spectral CL + UV	<b>87.8</b> $\pm$ 0.3	71.7 $\pm$ 0.5	<b>76.6</b> $\pm$ 0.6	<b>41.0</b> $\pm$ 6.4	29.3 $\pm$ 4.8	21.5 $\pm$ 6.2
SimSiam	<b>40.0</b> $\pm$ 0.5	<b>39.9</b> $\pm$ 0.3	<b>39.6</b> $\pm$ 0.3	16.9 $\pm$ 0.3	16.6 $\pm$ 0.4	16.9 $\pm$ 0.4
SimSiam + UV	35.4 $\pm$ 0.4	35.4 $\pm$ 0.2	34.5 $\pm$ 0.3	16.5 $\pm$ 0.2	16.5 $\pm$ 0.3	16.3 $\pm$ 0.5
Supervised	89.6 $\pm$ 0.1	78.3 $\pm$ 0.4	76.3 $\pm$ 1.1	59.2 $\pm$ 0.2	47.9 $\pm$ 0.2	43.9 $\pm$ 0.3

# Experimental results

## Results in the unsupervised case

Method	CIFAR 10			CIFAR 100		
	Label skew	Covariate shift	Joint shift	Label skew	Covariate shift	Joint shift
Local SimCLR	79.4 $\pm$ 0.2	<b>74.3</b> $\pm$ 0.3	71.0 $\pm$ 0.4	42.2 $\pm$ 0.2	<b>41.2</b> $\pm$ 0.2	38.1 $\pm$ 0.3
Federated SimCLR	<b>85.0</b> $\pm$ 0.2	73.8 $\pm$ 0.2	<b>74.8</b> $\pm$ 0.5	<b>48.5</b> $\pm$ 0.1	39.5 $\pm$ 0.2	<b>43.1</b> $\pm$ 0.2
Spectral CL	76.5 $\pm$ 1.1	<b>73.5</b> $\pm$ 0.4	68.2 $\pm$ 0.6	33.3 $\pm$ 6.0	<b>33.6</b> $\pm$ 2.3	<b>29.6</b> $\pm$ 6.2
Spectral CL + UV	<b>87.8</b> $\pm$ 0.3	71.7 $\pm$ 0.5	<b>76.6</b> $\pm$ 0.6	<b>41.0</b> $\pm$ 6.4	29.3 $\pm$ 4.8	21.5 $\pm$ 6.2
SimSiam	<b>40.0</b> $\pm$ 0.5	<b>39.9</b> $\pm$ 0.3	<b>39.6</b> $\pm$ 0.3	16.9 $\pm$ 0.3	16.6 $\pm$ 0.4	16.9 $\pm$ 0.4
SimSiam + UV	35.4 $\pm$ 0.4	35.4 $\pm$ 0.2	34.5 $\pm$ 0.3	16.5 $\pm$ 0.2	16.5 $\pm$ 0.3	16.3 $\pm$ 0.5
Supervised	89.6 $\pm$ 0.1	78.3 $\pm$ 0.4	76.3 $\pm$ 1.1	59.2 $\pm$ 0.2	47.9 $\pm$ 0.2	43.9 $\pm$ 0.3

## Results in the semi-supervised case

Method	CIFAR 10			CIFAR 100		
	Label skew	Covariate shift	Joint shift	Label Skew	Covariate shift	Joint shift
Local SimCLR	74.5 $\pm$ 0.3	<b>49.1</b> $\pm$ 1.3	45.8 $\pm$ 1.4	30.3 $\pm$ 0.2	15.1 $\pm$ 0.4	13.1 $\pm$ 0.3
Federated SimCLR	<b>78.0</b> $\pm$ 0.2	<b>50.3</b> $\pm$ 1.1	<b>49.9</b> $\pm$ 1.4	<b>34.5</b> $\pm$ 0.3	14.8 $\pm$ 0.3	<b>14.6</b> $\pm$ 0.3
Spectral CL	74.2 $\pm$ 0.3	48.0 $\pm$ 0.7	45.4 $\pm$ 1.5	30.1 $\pm$ 0.2	14.1 $\pm$ 0.4	12.3 $\pm$ 0.3
Spectral CL + UV	<b>79.6</b> $\pm$ 0.3	<b>49.7</b> $\pm$ 1.0	<b>49.8</b> $\pm$ 1.1	<b>34.0</b> $\pm$ 0.2	13.7 $\pm$ 0.3	<b>13.6</b> $\pm$ 0.4
SimSiam	75.3 $\pm$ 0.4	46.8 $\pm$ 0.7	40.5 $\pm$ 0.9	30.7 $\pm$ 0.2	13.4 $\pm$ 0.3	12.8 $\pm$ 0.3
SimSiam + UV	<b>80.4</b> $\pm$ 0.2	<b>50.0</b> $\pm$ 1.2	<b>44.3</b> $\pm$ 1.0	<b>34.3</b> $\pm$ 0.1	13.6 $\pm$ 0.3	<b>14.0</b> $\pm$ 0.4
Supervised	75.1 $\pm$ 0.2	48.1 $\pm$ 0.9	42.7 $\pm$ 1.7	29.6 $\pm$ 0.3	12.6 $\pm$ 0.2	12.2 $\pm$ 0.1

# Thank you

**Qualcomm**

Follow us on: [in](#) [X](#) [@](#) [▶](#) [f](#)

For more information, visit us at:

[qualcomm.com](http://qualcomm.com) & [qualcomm.com/blog](http://qualcomm.com/blog)

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

© Qualcomm Technologies, Inc. and/or its affiliated companies. All Rights Reserved.

Qualcomm is a trademark or registered trademark of Qualcomm Incorporated. Other products and brand names may be trademarks or registered trademarks of their respective owners.

References in this presentation to "Qualcomm" may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes our licensing business, QTL, and the vast majority of our patent portfolio. Qualcomm Technologies, Inc., a subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of our engineering, research and development functions, and substantially all of our products and services businesses, including our QCT semiconductor business.

Snapdragon and Qualcomm branded products are products of Qualcomm Technologies, Inc. and/or its subsidiaries. Qualcomm patented technologies are licensed by Qualcomm Incorporated.