# Calibrated and Communication Efficient Federated Learning

AAAI Workshop on Edge Intelligence, Vancouver (2024-02-26)

Pascal Poupart, CIFAR AI Chair Vector Institute
David R. Cheriton School of Computer Science
University of Waterloo

UNIVERSITY OF WATERLOO

VECTOR INSTITUTE

UNIVERSITY OF WATERLOO

VECTOR INSTITUTE

# Research topics

- **Machine Learning:** reinforcement learning, uncertainty quantification, federated learning, inverse constraint learning
- **Natural Language Processing:** knowledge graphs, post-editing ASR error correction, conversational agents
- **Applications**: autonomous driving, sports analytics, material design for CO2 recycling

# Industry Partners

intact

The Schulich Foundation

Microsoft

Google

Ford

SPORT LOGiQ

Manulife

kik

InTheChat

HOCKEYTECH

ELEMENT AI
consultant

BMO

SCRIBENDI

HUAWEI
consultant

focal Systems

PRONAV Technologies Ltd
consultant

RBC
research director

BOREALIS AI

TalkIQ
consultant

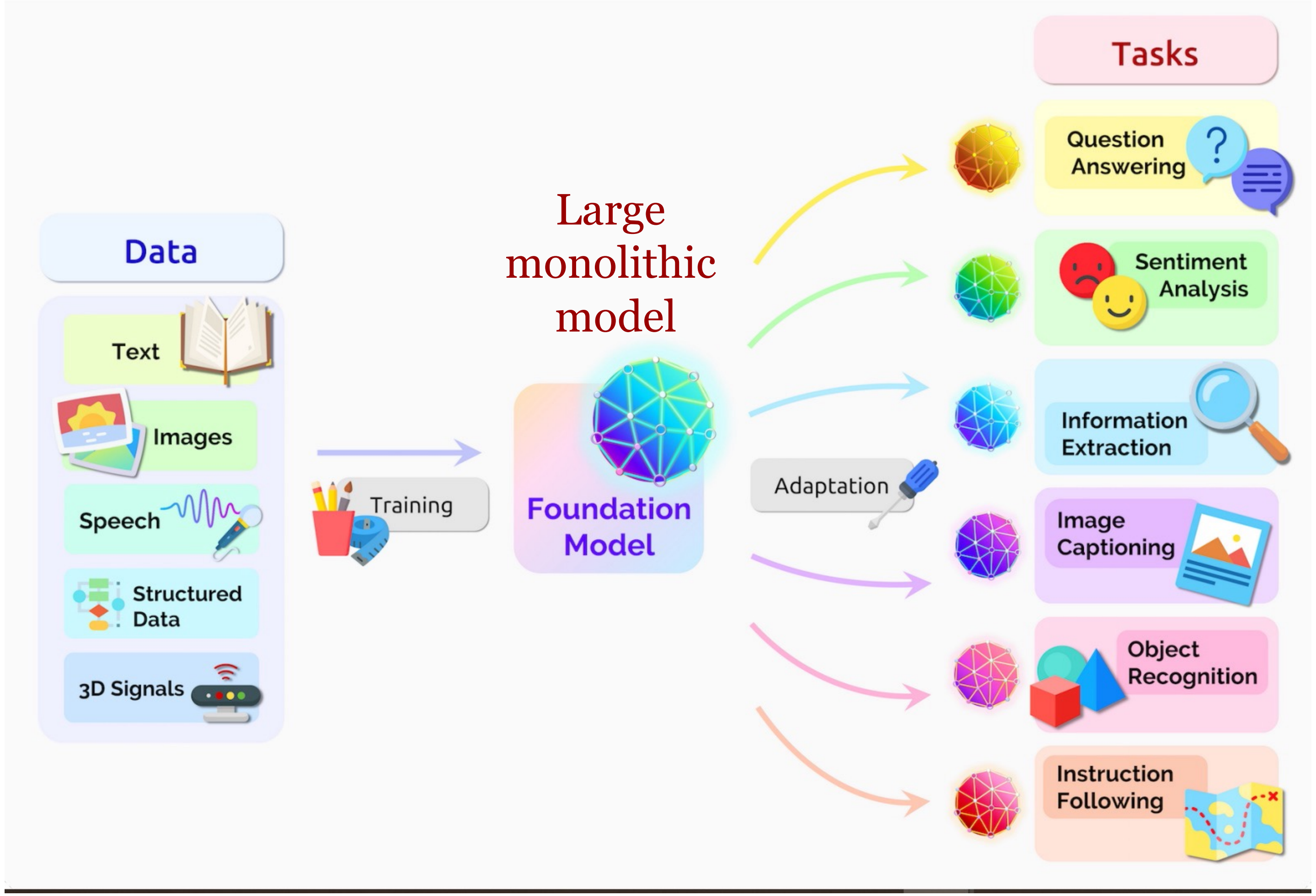
UNIVERSITY OF WATERLOO

# Outline

- Federated Learning Background
- Calibrated One-Round Federated Learning
  - Mohsin Hasan, Guojung Zhang, Kaiyang Guo, Xi Chen, Pascal Poupart, **Calibrated One Round Federated Learning with Bayesian Inference in the Predictive Space**, AAAI, 2024.
- Flexible and Communication Efficient Federated Learning
  - Haolin Yu, Guojun Zhang, Pascal Poupart, **FedLog: Federated Classification with Less Communication and More Flexibility**, under review.
- RL Foundation Models

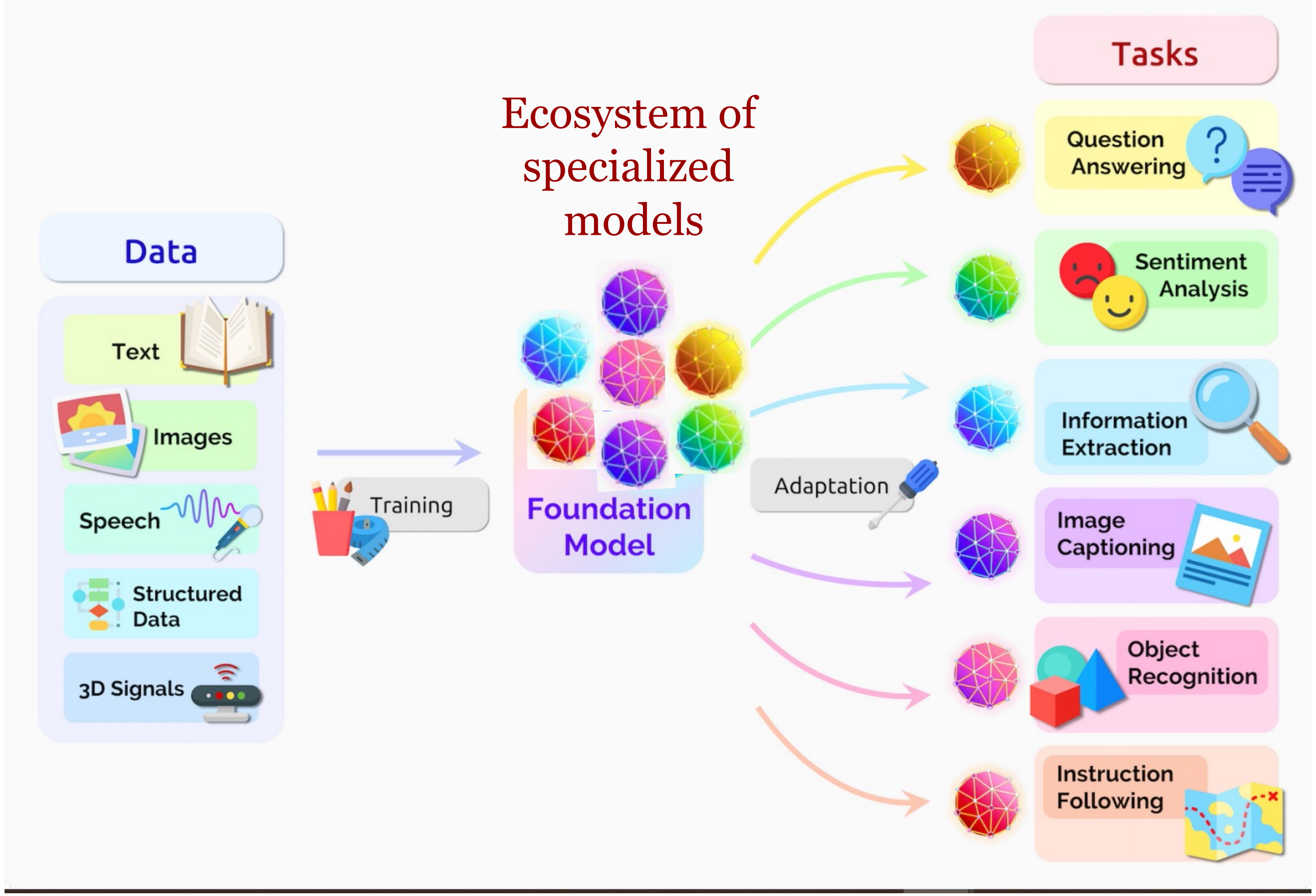
# Foundation Models

Credit: On the Opportunities and Risks of Foundation Models (2022)



# Foundation Models

Credit: On the Opportunities and Risks of Foundation Models (2022)




VECTOR INSTITUTE

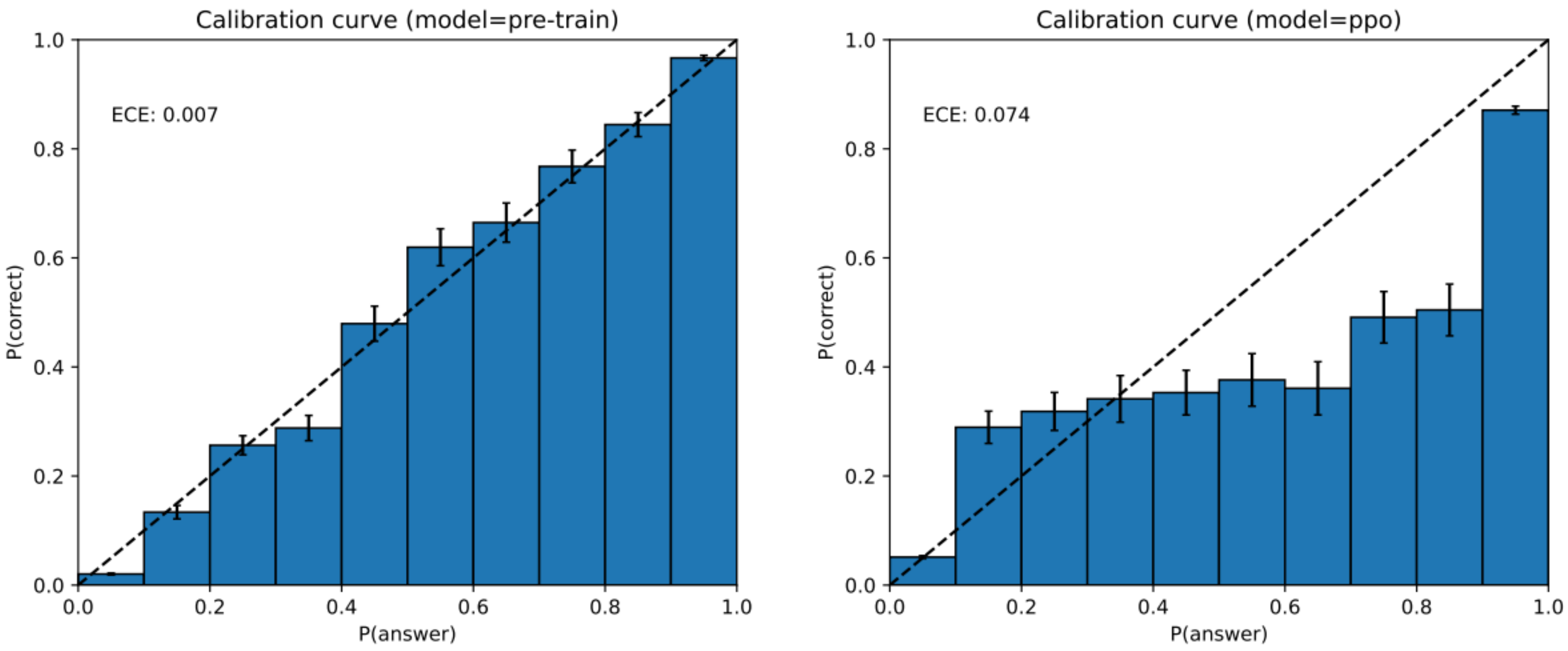


UNIVERSITY OF WATERLOO

# GPT-4 Calibration

- Expected Calibration Error: $ECE = \sum_m \frac{|B_m|}{n} |acc(B_m) - conf(B_m)|$

Credit: GPT-4 Technical Report (OpenAI)



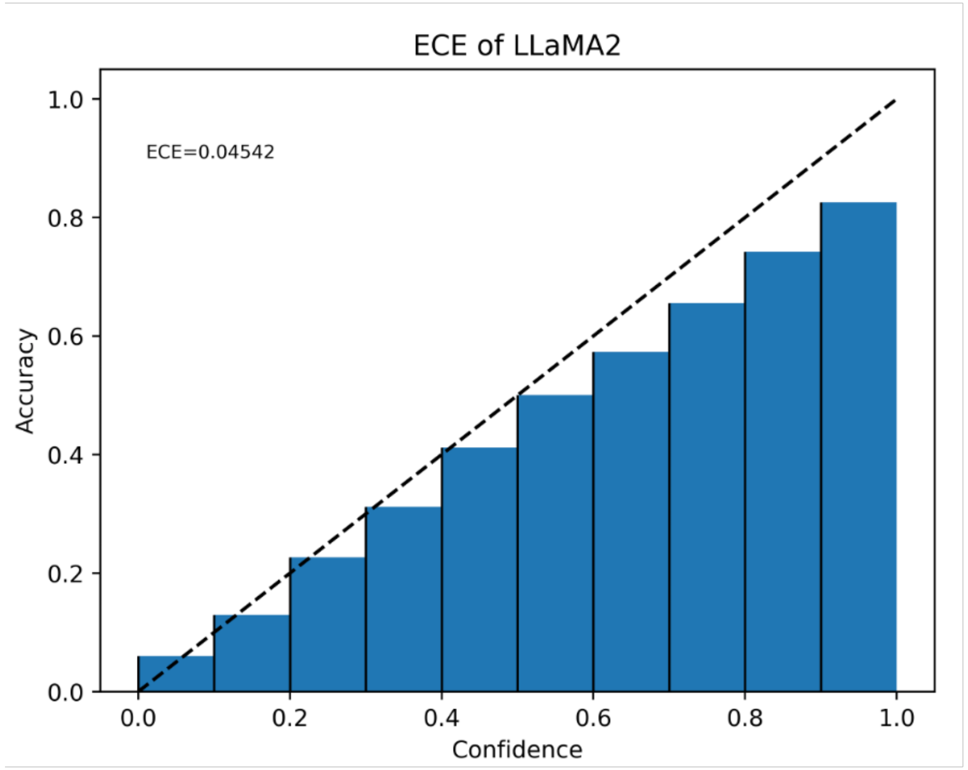

VECTOR INSTITUTE



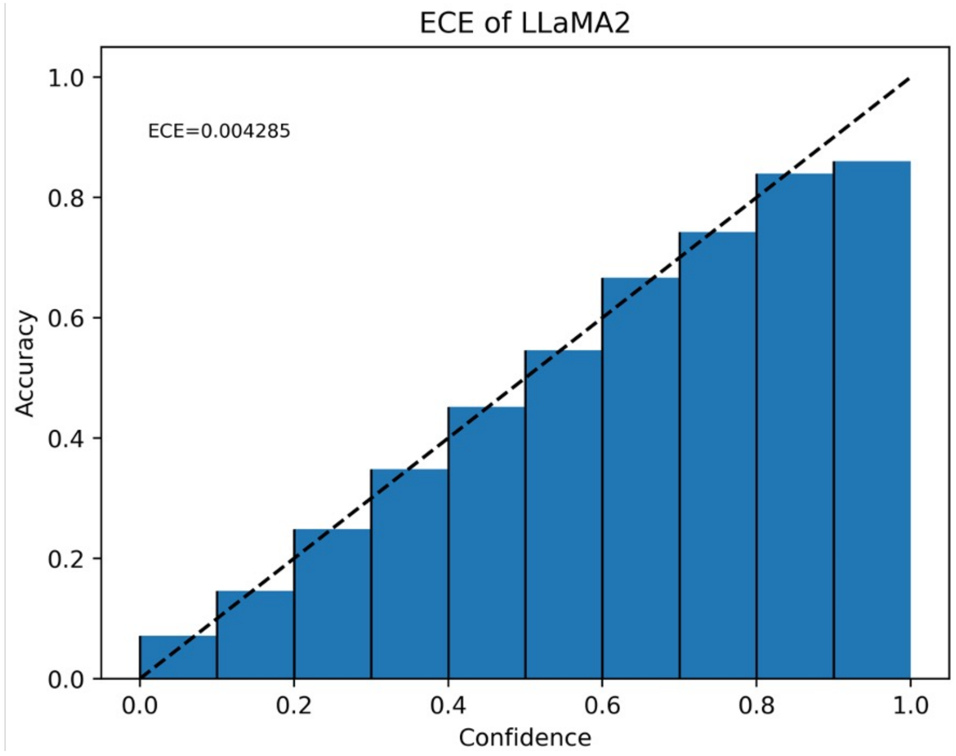
UNIVERSITY OF WATERLOO

# LLAMA-2 Calibration

Pre-trained model




Bin-wise temperature scaling



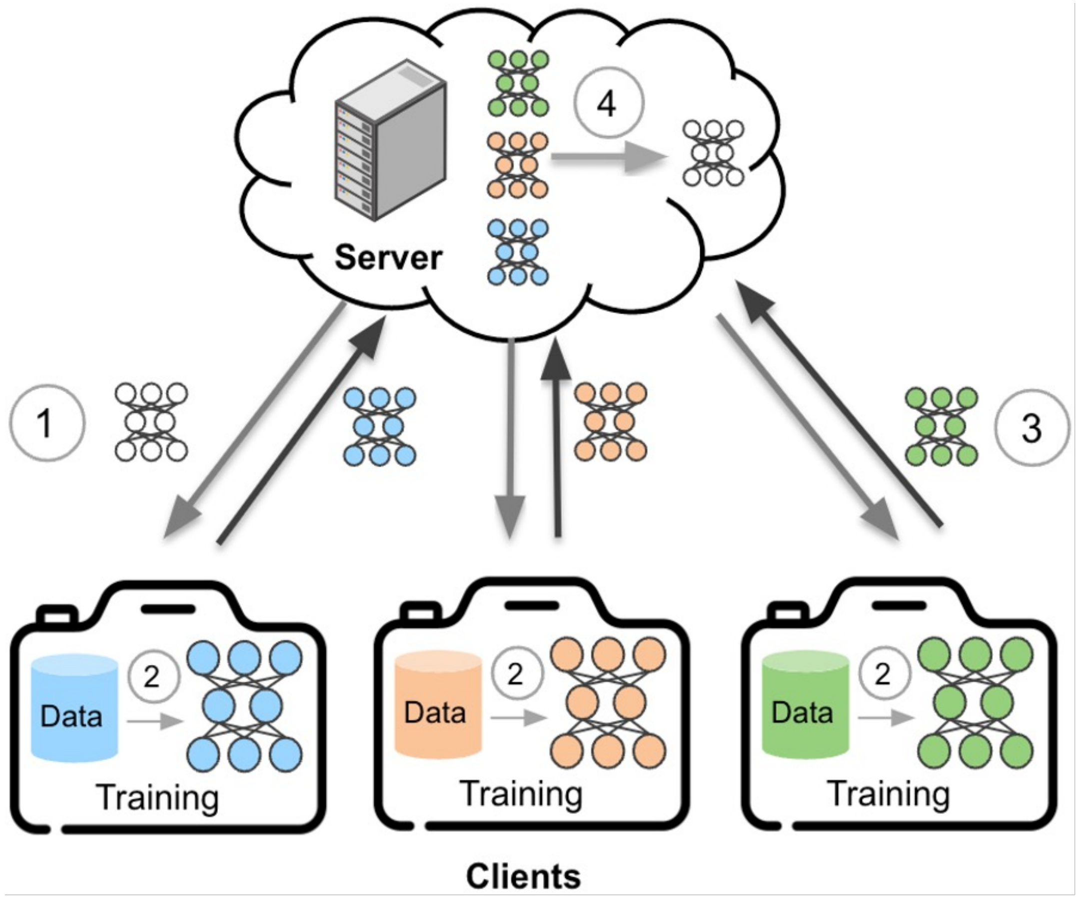

Ruotian Wu and Ahmad Rashid

VECTOR INSTITUTE

UNIVERSITY OF WATERLOO

# Federated Learning

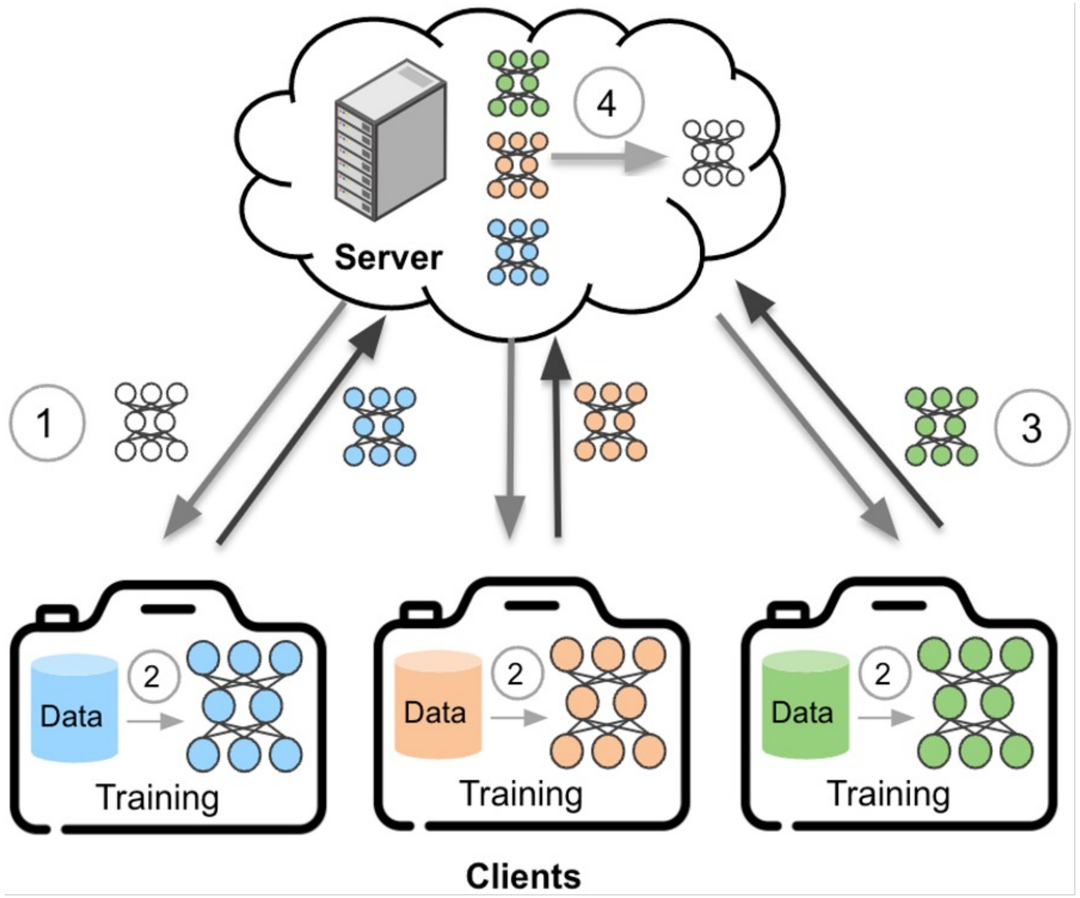
- Communication cost:
  - # of communication rounds
  - Message sizes
- Calibration?




Credit: https://ai.sony/blog/blog-032/

# Fed-Averaging

- Costly communication:
  - 1000's of rounds
  - Message size: #params
- Often mis-calibrated




Credit: https://ai.sony/blog/blog-032/

# Outline

- Federated Learning Background
- Calibrated One-Round Federated Learning
  - Mohsin Hasan, Guojung Zhang, Kaiyang Guo, Xi Chen, Pascal Poupart, **Calibrated One Round Federated Learning with Bayesian Inference in the Predictive Space**, AAAI, 2024.
- Flexible and Communication Efficient Federated Learning
  - Haolin Yu, Guojun Zhang, Pascal Poupart, **FedLog: Federated Classification with Less Communication and More Flexibility**, under review.
- RL Foundation Models

# Bayesian Federated Learning

- Bayes Theorem: $P(\theta|D_1, D_2, \ldots, D_n) = k\, P(\theta)P(D_1|\theta)P(D_2|\theta) \ldots P(D_n|\theta)$

$$= \underbrace{\frac{k}{P(\theta)^{n-1}}}_{\text{server}} \underbrace{P(\theta|D_1)}_{\text{client 1}} \underbrace{P(\theta|D_2)}_{\text{client 2}} \ldots \underbrace{P(\theta|D_n)}_{\text{client n}}$$

- **Single communication round**
- **Well calibrated**

- Catch: intractable computation
  - Parameter posterior: $P(\theta|D_1, D_2, \ldots, D_n)$
  - Predictive posterior: $\mathrm{P}(y|x, D_1, D_2, \ldots, D_n) = \int_\theta P(y|x, \theta)\, P(\theta|D_1, D_2, \ldots, D_n)\, d\theta$

VECTOR INSTITUTE



UNIVERSITY OF WATERLOO

# Bayesian Federated Learning in Predictive Space

- Bayesian Committee machine
  - $P(y|x, D_1, D_2, \ldots, D_n) = k\, P(y|x) P(D_1|y, x) P(D_2|D_1, y, x) \ldots P(D_n|D_{n-1}, \ldots, D_1, y, x)$

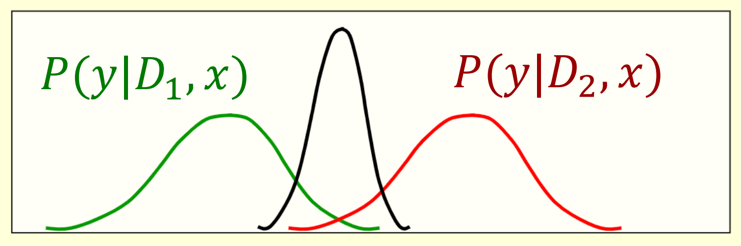
$$\approx k\, P(y|x) P(D_1|y, x) P(D_2|y, x) \ldots P(D_n|\, y, x)$$

$$= \underbrace{\frac{k}{P(y|x)^{n-1}}}_{\text{server}} \underbrace{P(y|D_1, x)}_{\text{client 1}} \underbrace{P(y|D_2, x)}_{\text{client 2}} \ldots \underbrace{P(y|D_n, x)}_{\text{client n}}$$

- Calibration?

| **Aggregation** | **Homogeneous (identical clients)** | **Heterogeneous (independent clients)** |
|---|---|---|
| Product of experts | | |

$P(y|D_1, D_2, x)$




Credit: Geoff Hinton

# Bayesian Federated Learning in Predictive Space

- Bayesian Committee machine
  - $P(y|x, D_1, D_2, \dots, D_n) = k\, P(y|x) P(D_1|y,x) P(D_2|D_1,y,x) \dots P(D_n|D_{n-1},\dots,D_1,y,x)$

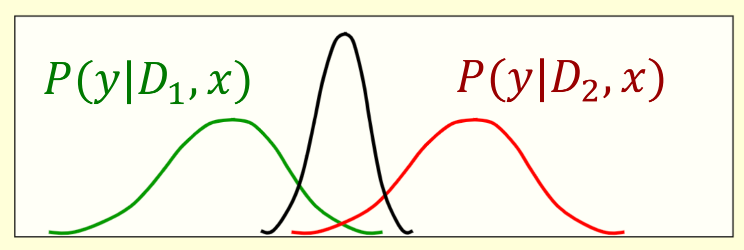
$$\approx k\, P(y|x) P(D_1|y,x) P(D_2|y,x) \dots P(D_n|\, y,x)$$

$$= \underbrace{\frac{k}{P(y|x)^{n-1}}}_{\text{server}} \underbrace{P(y|D_1,x)}_{\text{client 1}} \underbrace{P(y|D_2,x)}_{\text{client 2}} \dots \underbrace{P(y|D_n,x)}_{\text{client n}}$$

- Calibration?

| Aggregation | Homogeneous (identical clients) | Heterogeneous (independent clients) |
|---|---|---|
| Product of experts | overconfident | calibrated |

$P(y|D_1, D_2, x)$




Credit: Geoff Hinton

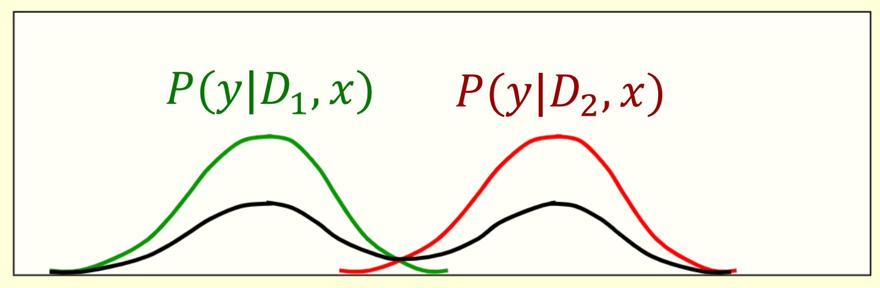
# Mixture of Experts

- $P(y|x, D_1, D_2, \dots, D_n) = \frac{|D_1|}{|D|} \underbrace{P(y|x, D_1)}_{\text{client 1}} + \frac{|D_2|}{|D|} \underbrace{P(y|x, D_2)}_{\text{client 2}} + \cdots + \frac{|D_n|}{|D|} \underbrace{P(y|x, D_n)}_{\text{client n}}$

- Calibration?

| Aggregation | Homogeneous (identical clients) | Heterogeneous (independent clients) |
|---|---|---|
| Product of experts | overconfident | calibrated |
| Mixture of experts | | |

Credit: Geoff Hinton




$P(y|D_1, D_2, x)$

VECTOR INSTITUTE



UNIVERSITY OF WATERLOO

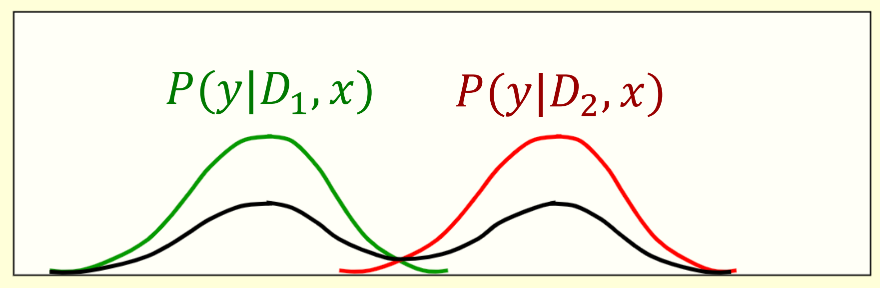
# Mixture of Experts

- $P(y|x, D_1, D_2, \ldots, D_n) = \frac{|D_1|}{|D|} \underbrace{P(y|x, D_1)}_{\text{client 1}} + \frac{|D_2|}{|D|} \underbrace{P(y|x, D_2)}_{\text{client 2}} + \cdots + \frac{|D_n|}{|D|} \underbrace{P(y|x, D_n)}_{\text{client n}}$

- Calibration?

| Aggregation | Homogeneous (identical clients) | Heterogeneous (independent clients) |
|---|---|---|
| Product of experts | overconfident | calibrated |
| Mixture of experts | calibrated | underconfident |

Credit: Geoff Hinton




$P(y|D_1, D_2, x)$

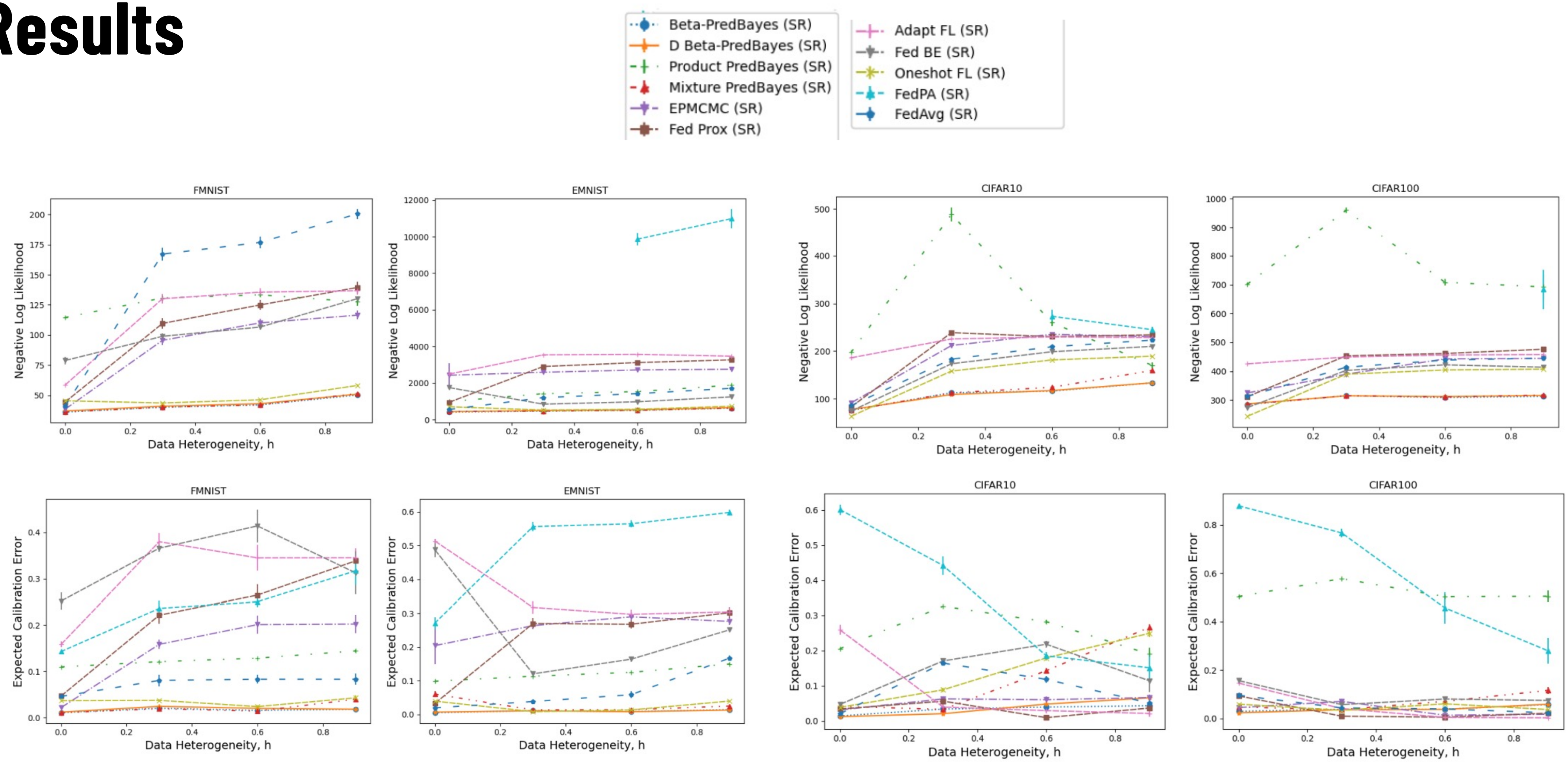
# $\beta$-PredBayes

- Interpolate between product and mixture of experts

- Objective: $\min_\beta \log P_\beta(y|x, D)$

where $\log P_\beta(y|x,D) = \underbrace{\beta \log\left(\frac{1}{P(y|x)^{n-1}} \prod_i P(y|x,D_i)\right)}_{\text{Product of experts}} + (1-\beta) \underbrace{\log\left(\sum_i \frac{|D_i|}{|D|} P(y|x,D_i)\right)}_{\text{Mixture of experts}}$

# Results

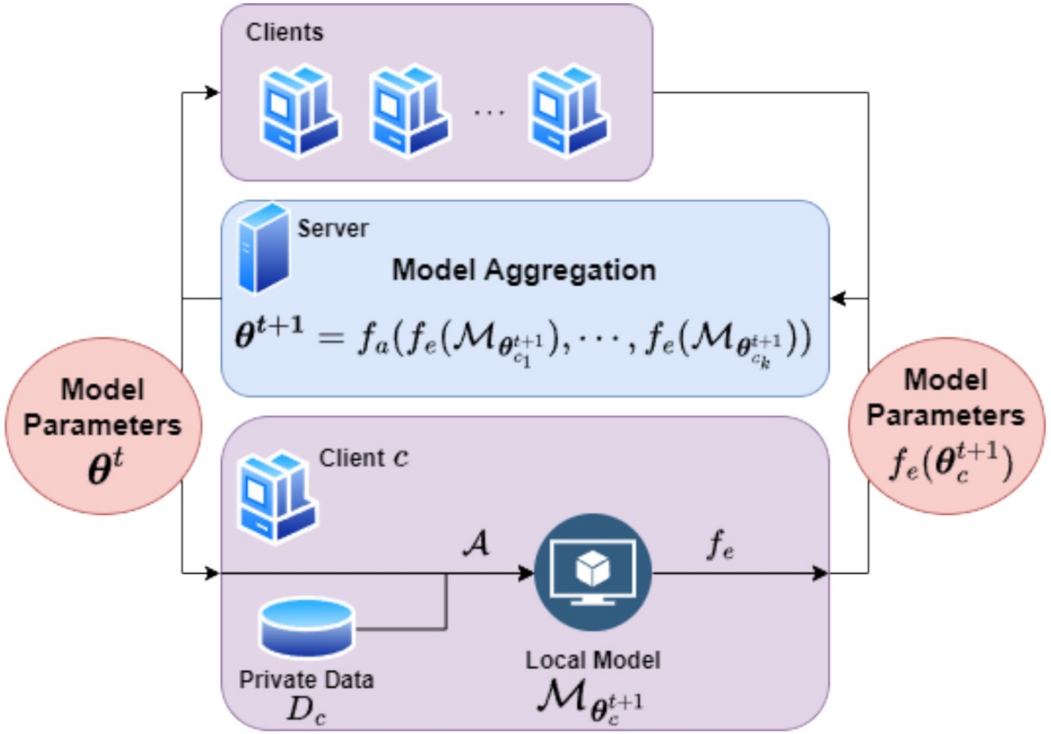

Beta-PredBayes (SR)
D Beta-PredBayes (SR)
Product PredBayes (SR)
Mixture PredBayes (SR)
EPMCMC (SR)
Fed Prox (SR)
Adapt FL (SR)
Fed BE (SR)
Oneshot FL (SR)
FedPA (SR)
FedAvg (SR)
FMNIST
EMNIST
CIFAR10
CIFAR100
Negative Log Likelihood
Expected Calibration Error
Data Heterogeneity, h

# Outline

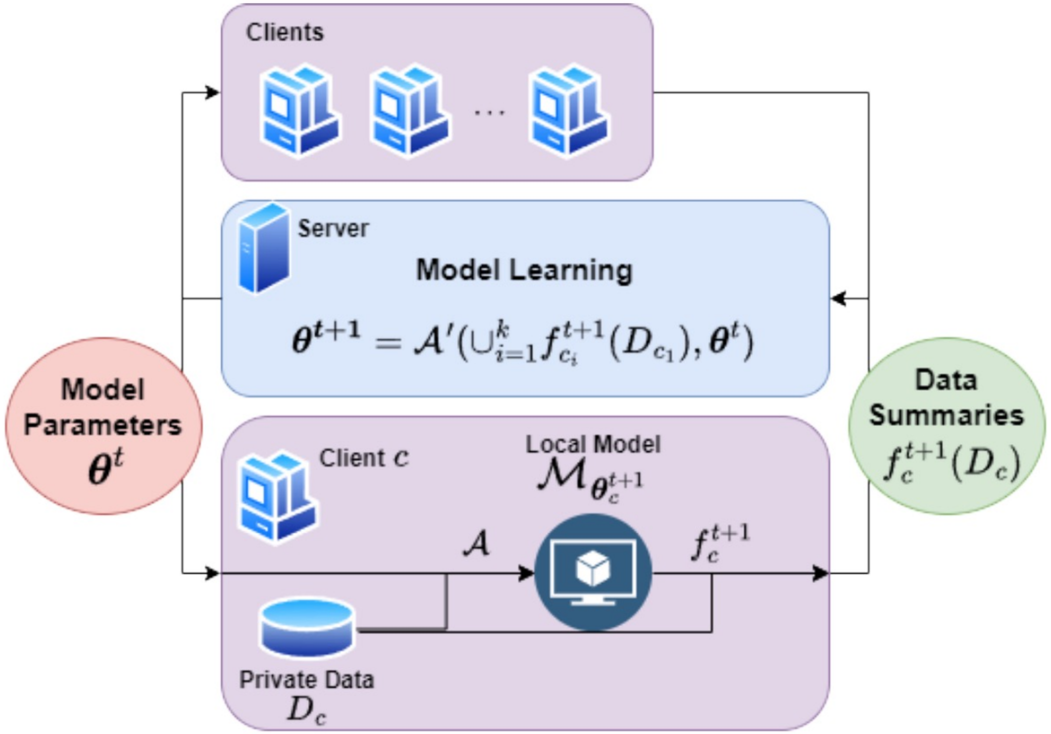
- Federated Learning Background
- Calibrated One-Round Federated Learning
  - Mohsin Hasan, Guojung Zhang, Kaiyang Guo, Xi Chen, Pascal Poupart, **Calibrated One Round Federated Learning with Bayesian Inference in the Predictive Space**, AAAI, 2024.
- Flexible and Communication Efficient Federated Learning
  - Haolin Yu, Guojun Zhang, Pascal Poupart, **FedLog: Federated Classification with Less Communication and More Flexibility**, under review.
- RL Foundation Models

# Federated Bayesian Logistic Regression

Traditional paradigm: share parameters




Alternative paradigm: share data summaries



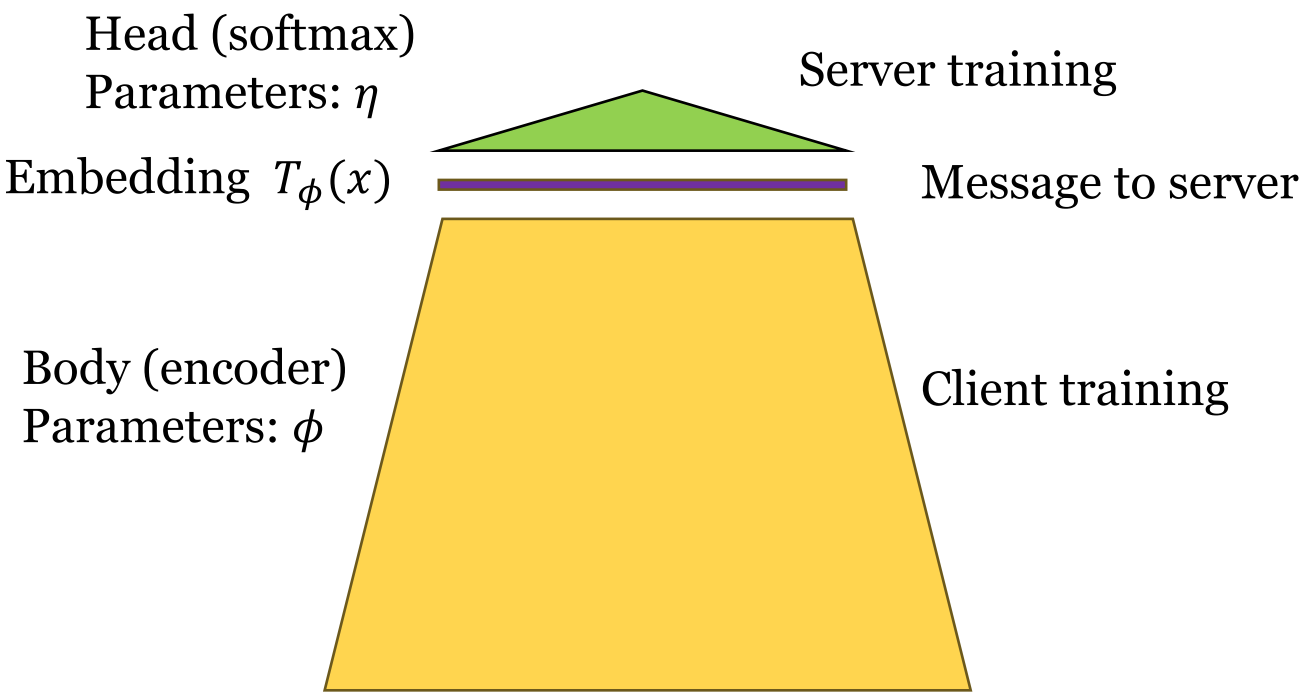

VECTOR INSTITUTE



UNIVERSITY OF WATERLOO

# Federated Bayesian Logistic Regression

- **Benefits:**
  - Smaller messages (embedding size)
  - Flexibility (clients can use different body architectures)



# Results

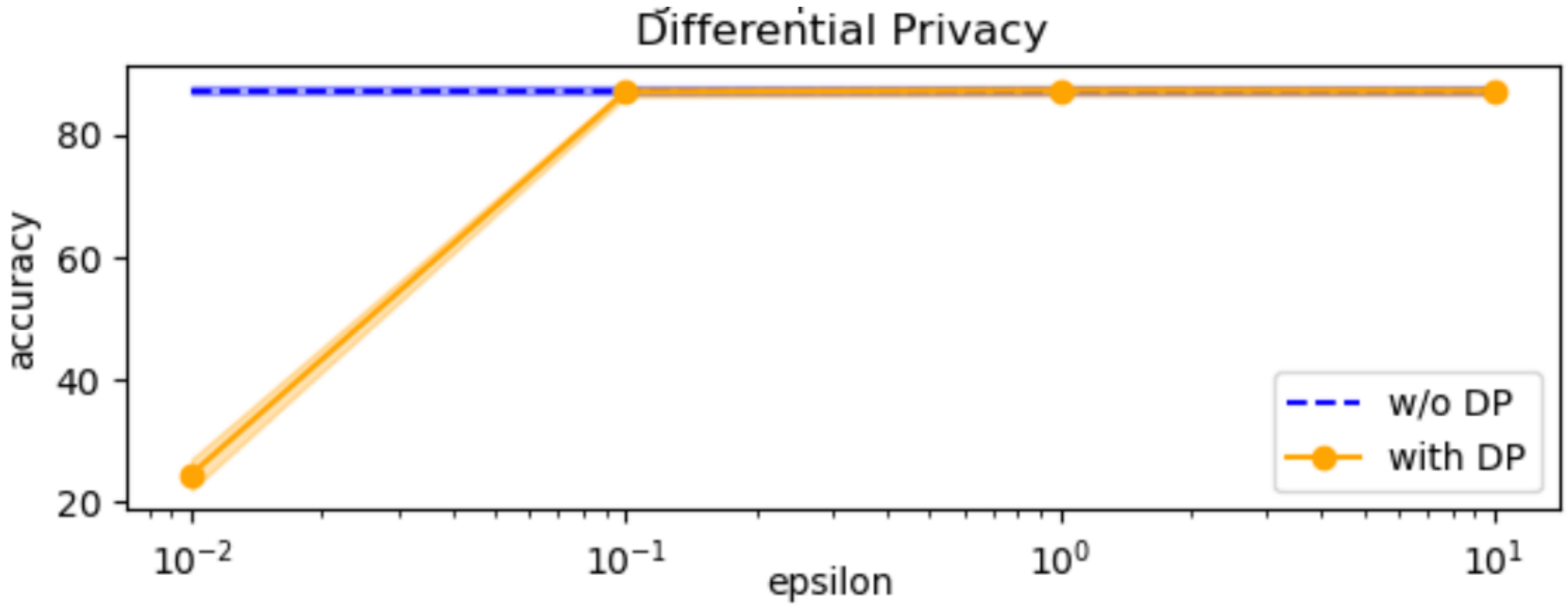
| | MNIST | | CIFAR10 | | CIFAR100 | |
|---|---|---|---|---|---|---|
| | accuracy | comm cost | accuracy | comm cost | accuracy | comm cost |
| FedAvg | 89.76±0.69⇓ | 698880 | 26.29±0.44⇓ | 4102528 | 13.34±0.15⇓ | 4684288 |
| LG-FedAvg 1 | 97.85±0.05⇓ | **16320** | 86.57±0.29⇓ | **64640** | 55.00±0.26⇓ | **646400** |
| LG-FedAvg 2 | **98.18±0.06** | 529920 | 85.56±0.32⇓ | 839040 | 54.90±0.24⇓ | 1420800 |
| FedPer | 96.16±0.19⇓ | 168960 | 83.54±0.40⇓ | 3263488 | 52.82±0.21⇓ | 3263488 |
| FedRep | 95.51±0.29⇓ | 168960 | 82.96±0.35⇓ | 3263488 | 48.70±0.29⇓ | 3263488 |
| CS-FL | 79.65±1.22⇓ | 72088 | 23.60±1.08⇓ | 423092 | 4.52±0.15⇓ | 483086 |
| **FedLog (ours)** | **98.15±0.05** | **16320** | **87.08±0.22** | **64640** | **56.46±0.27** | **646400** |

VECTOR INSTITUTE



UNIVERSITY OF WATERLOO

# Differential Privacy

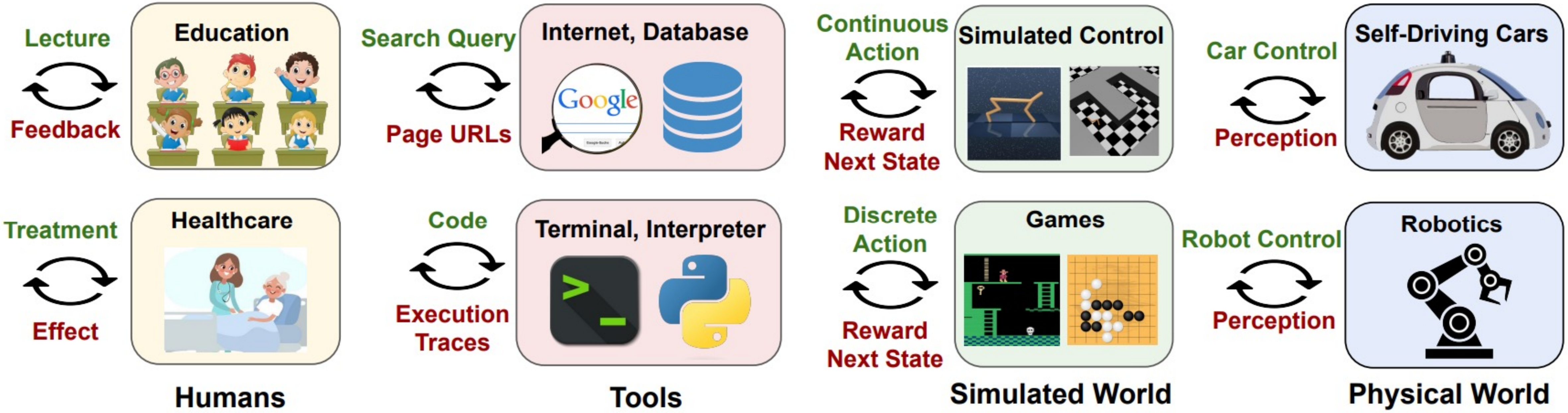


# Outline

- Federated Learning Background
- Calibrated One-Round Federated Learning
  - Mohsin Hasan, Guojung Zhang, Kaiyang Guo, Xi Chen, Pascal Poupart, **Calibrated One Round Federated Learning with Bayesian Inference in the Predictive Space**, AAAI, 2024.
- Flexible and Communication Efficient Federated Learning
  - Haolin Yu, Guojun Zhang, Pascal Poupart, **FedLog: Federated Classification with Less Communication and More Flexibility**, under review.
- RL Foundation Models

# Foundation Models as Interactive Systems

Yang, Nachum, Du, Wei, Abbeel, Schuurmans, Large Foundation Models for Decision Making: Problems, Methods, and Opportunities




VECTOR INSTITUTE



UNIVERSITY OF WATERLOO

# Vector RL Foundation Model

- Reduce data complexity with pre-trained RL Policy/Q-function
- Generalize across various business problems
- Evaluation partners

# Conclusion

- Distributed Foundation models
- Calibrated and Communication Efficient Federated Learning
  - Mohsin Hasan, Guojung Zhang, Kaiyang Guo, Xi Chen, Pascal Poupart, **Calibrated One Round Federated Learning with Bayesian Inference in the Predictive Space**, AAAI, 2024.
  - Haolin Yu, Guojun Zhang, Pascal Poupart, **FedLog: Federated Classification with Less Communication and More Flexibility**, under review.

- **Future work**
  - **RL Foundation Models**
  - **Composition of specialized models**