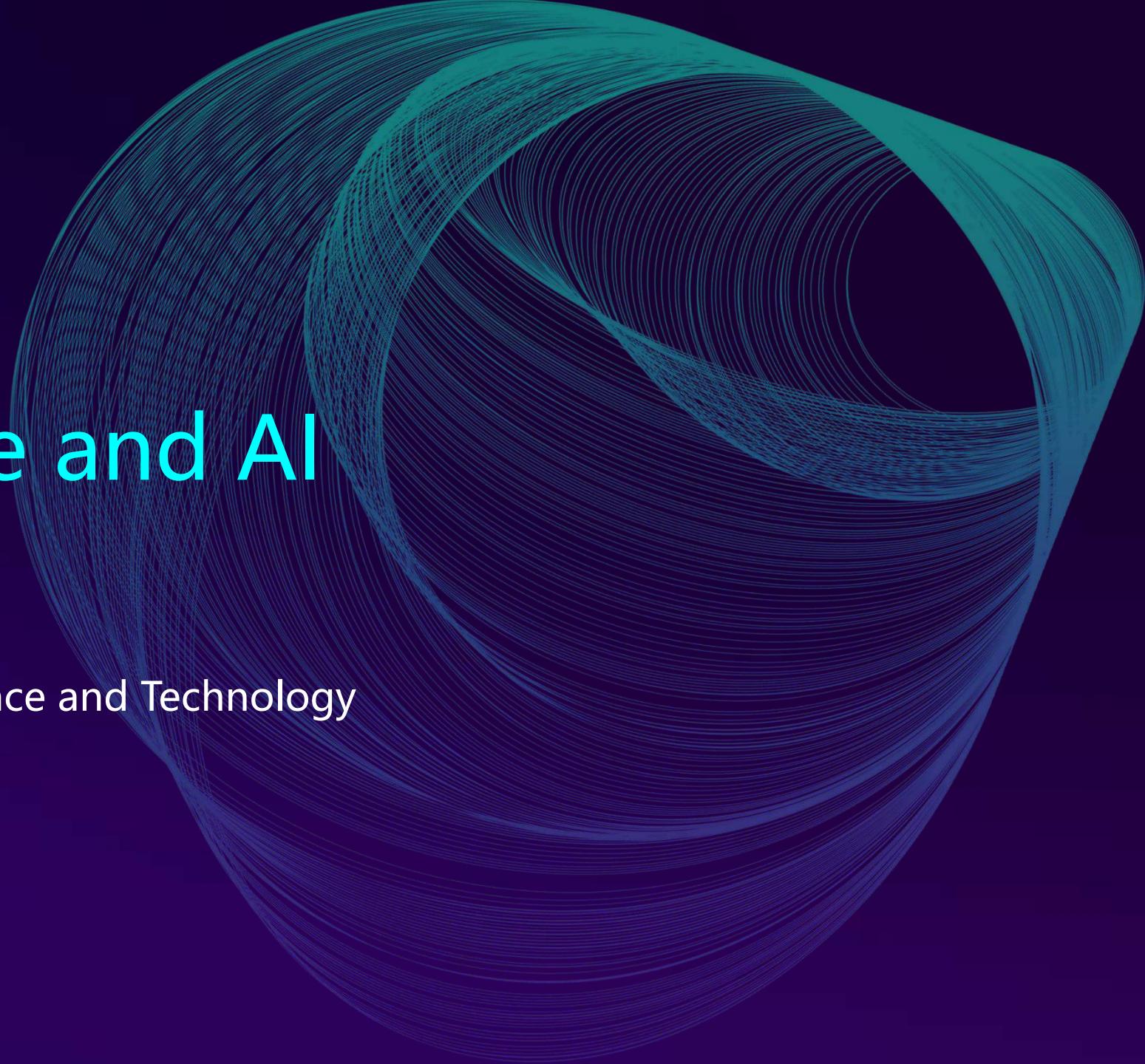


# **GDPR**

## Data Shortage and AI

**Qiang Yang**

Hong Kong University of Science and Technology



# AI and Big Data

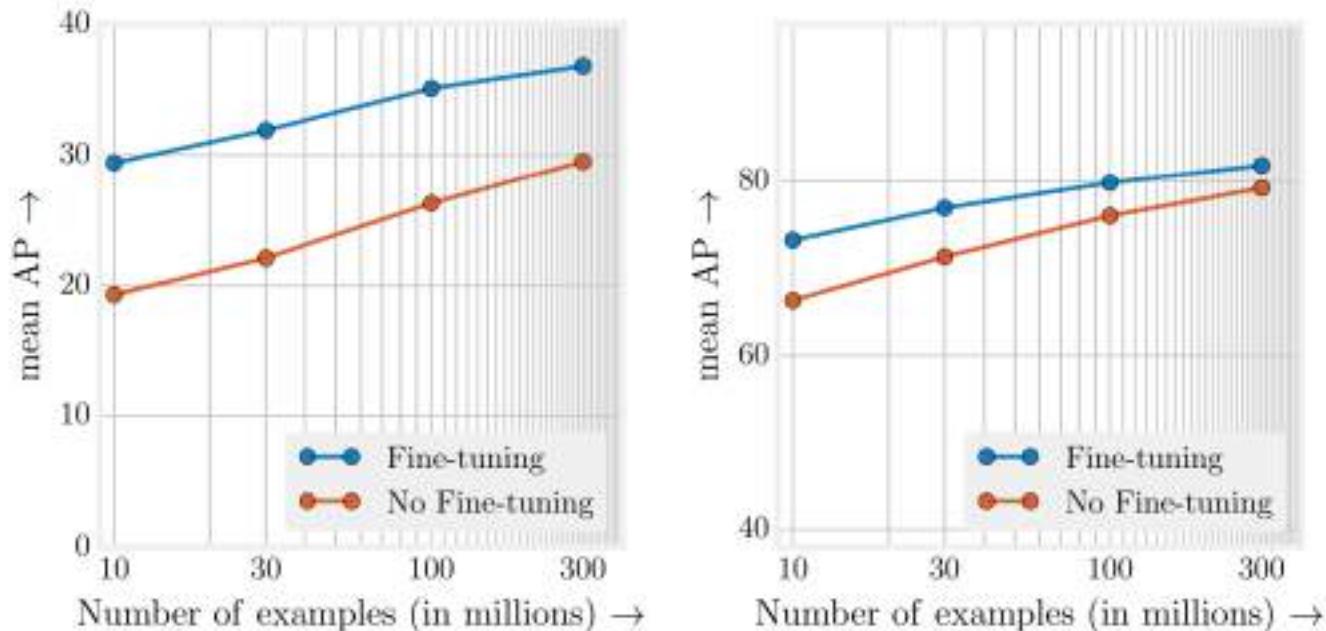


Figure 4. Object detection performance when initial checkpoints are pre-trained on different subsets of JFT-300M from scratch. x-axis is the data size in log-scale, y-axis is the detection performance in mAP@[.5,.95] on COCO minival\* (left), and in mAP@.5 on PASCAL VOC 2007 test (right).

# 1. Most Applications Have Only Small Data

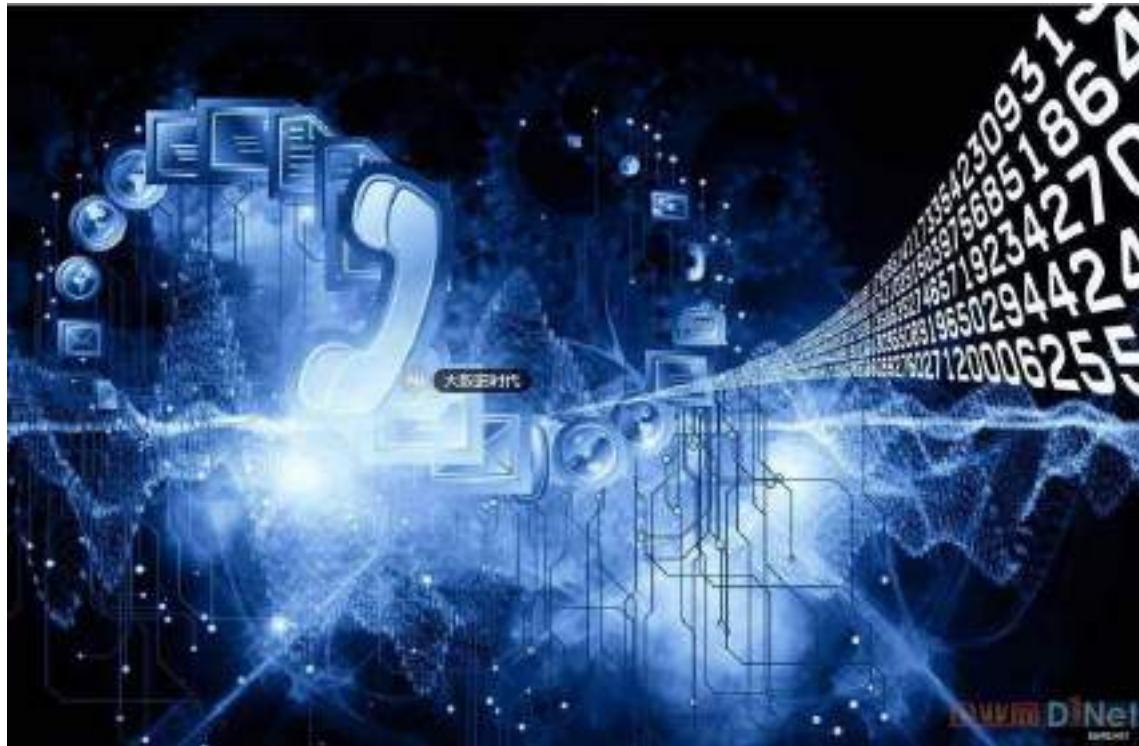
- Contract review law firms typically have annotated 10K - 20K of labeled contracts as samples (Bradley Arsenault, Electric Brain 2018)
- In finance industry, large loans are few, with only ~ 100 examples as typical samples (4paradigm.com, 2017)
- In medical image recognition, high-quality labeled data are few (A Survey on Deep Learning in Medical Image Analysis, Geert Litjens, et al. 2017 Arxiv.)

## 2. Data Sharing Among Parties: Difficult, Impossible or Immoral

- Medical clinical trial data cannot be shared (by Rogier Stegeman 2018 , on Genemetics)
- Our society demands more control on data privacy and security
  - GDPR, Government Regulations
  - Corporate Security and Confidentiality Concerns
  - Data privacy concerns



# Reality: Data often in form of Isolated Islands



# Two Challenges and Two Solutions

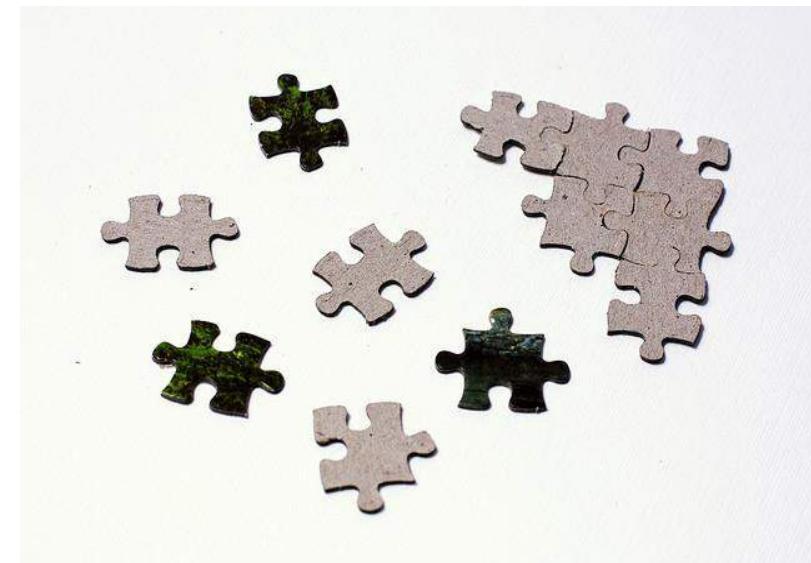
- **Small Data**

**Transfer Learning** from source data and models



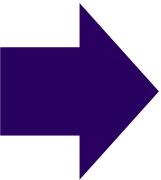
- **Fragmented Data**

**Federated learning** with many parties

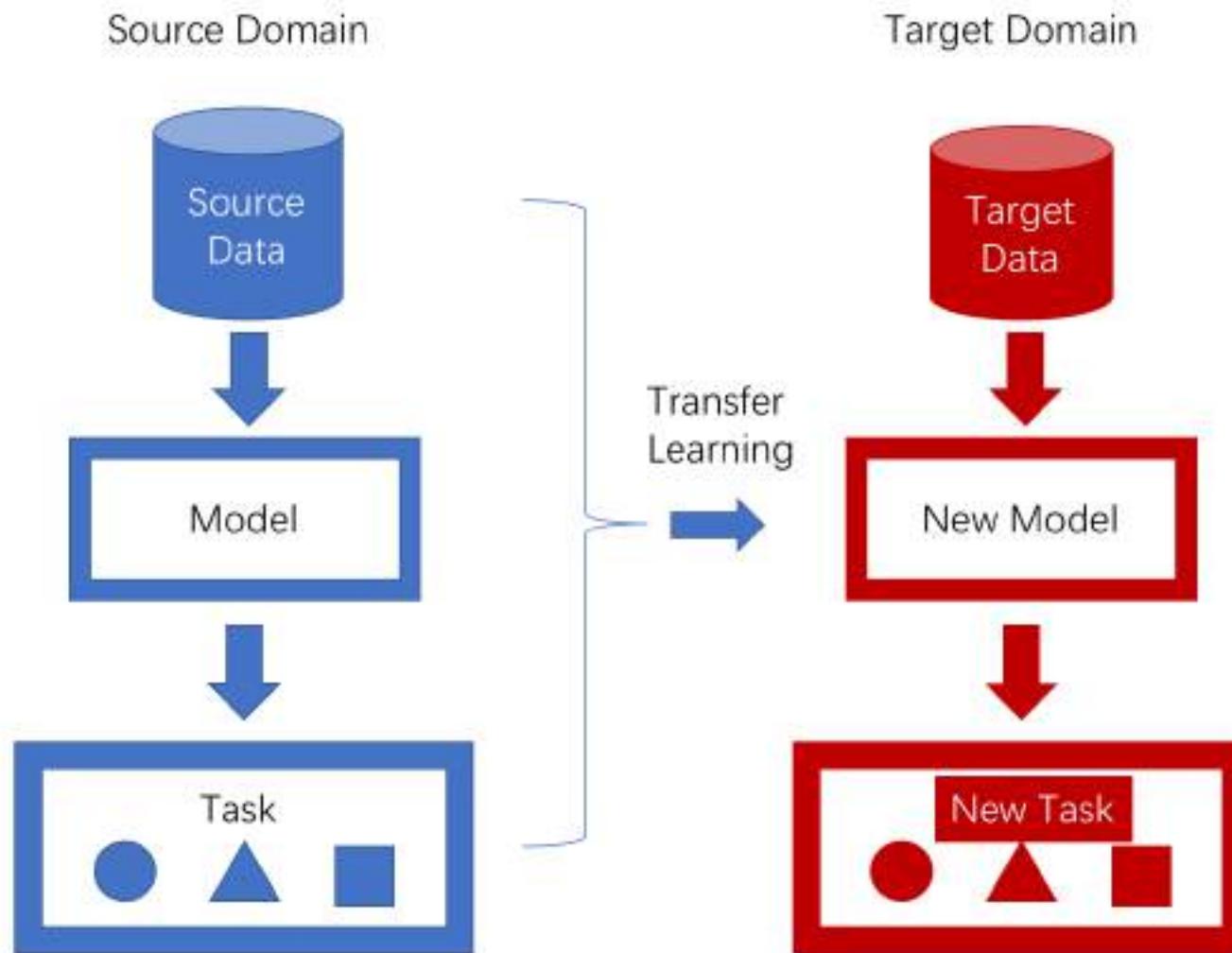


Often, these two problems occur together

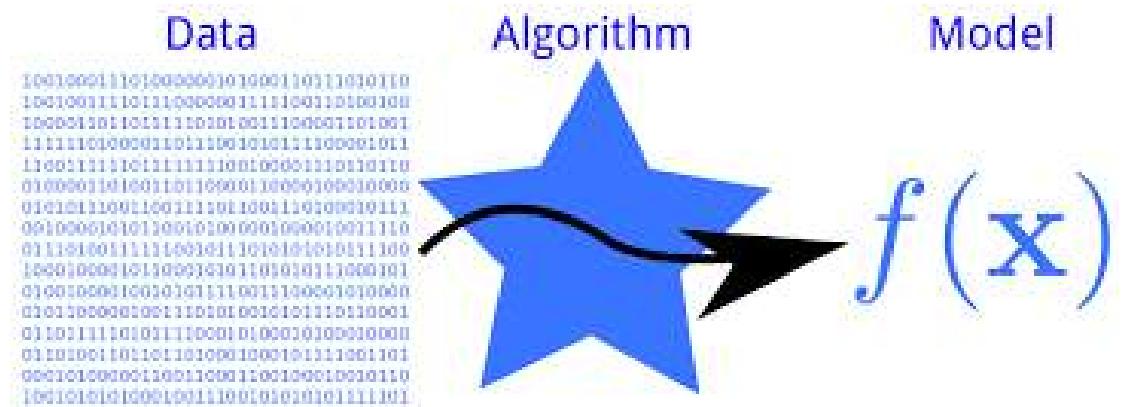
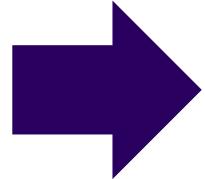
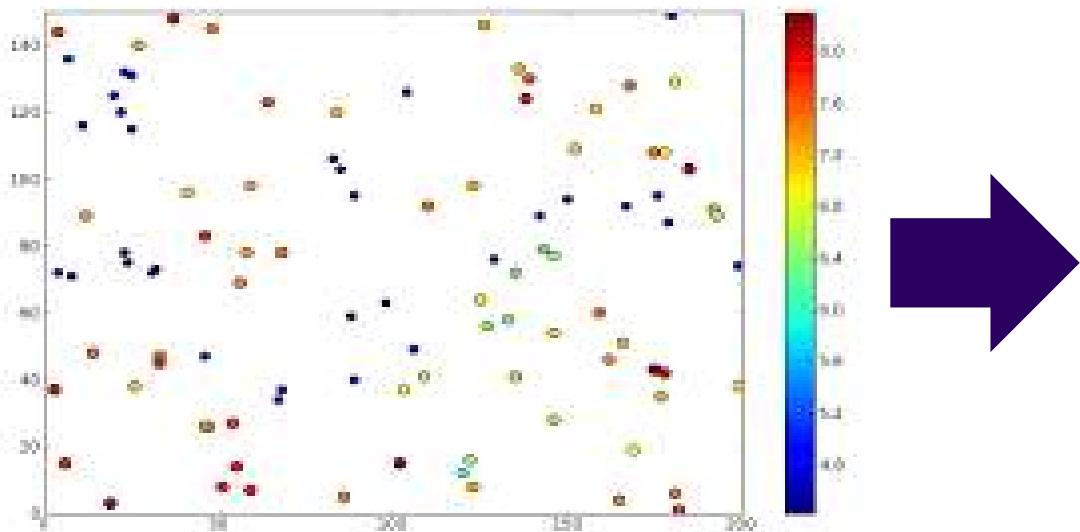
# Transfer Learning



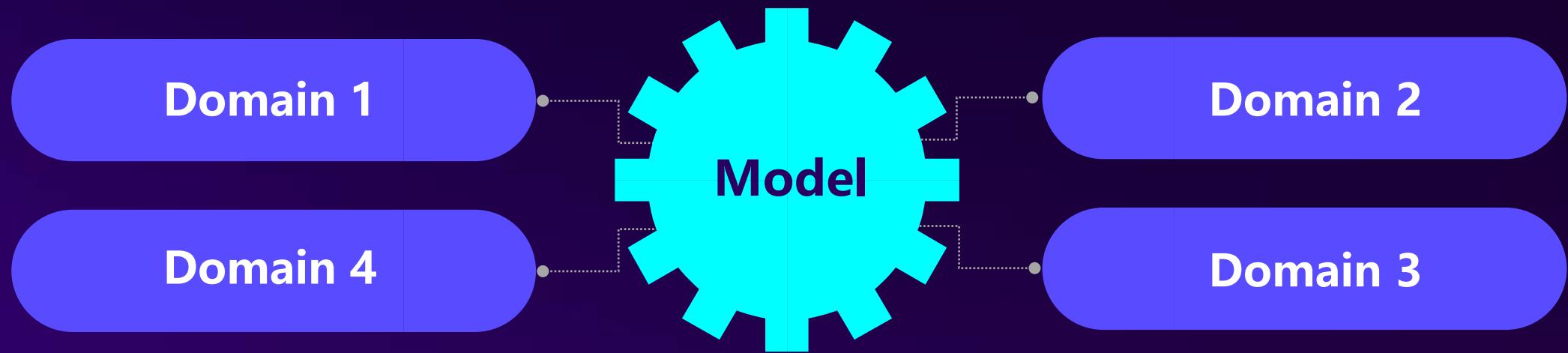
# Transfer Learning Models



# Why Transfer Learning? Small Data



# Why Transfer Learning : Reliability



# Why Transfer Learning? Personalization



# Learning to Transfer

## Research issues

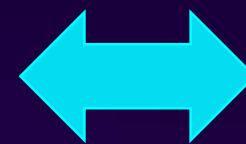
- When to transfer
- How to transfer
- What to transfer
- Learning how to learn by transfer learning



# Key to Transfer Learning : Finding the Invariance



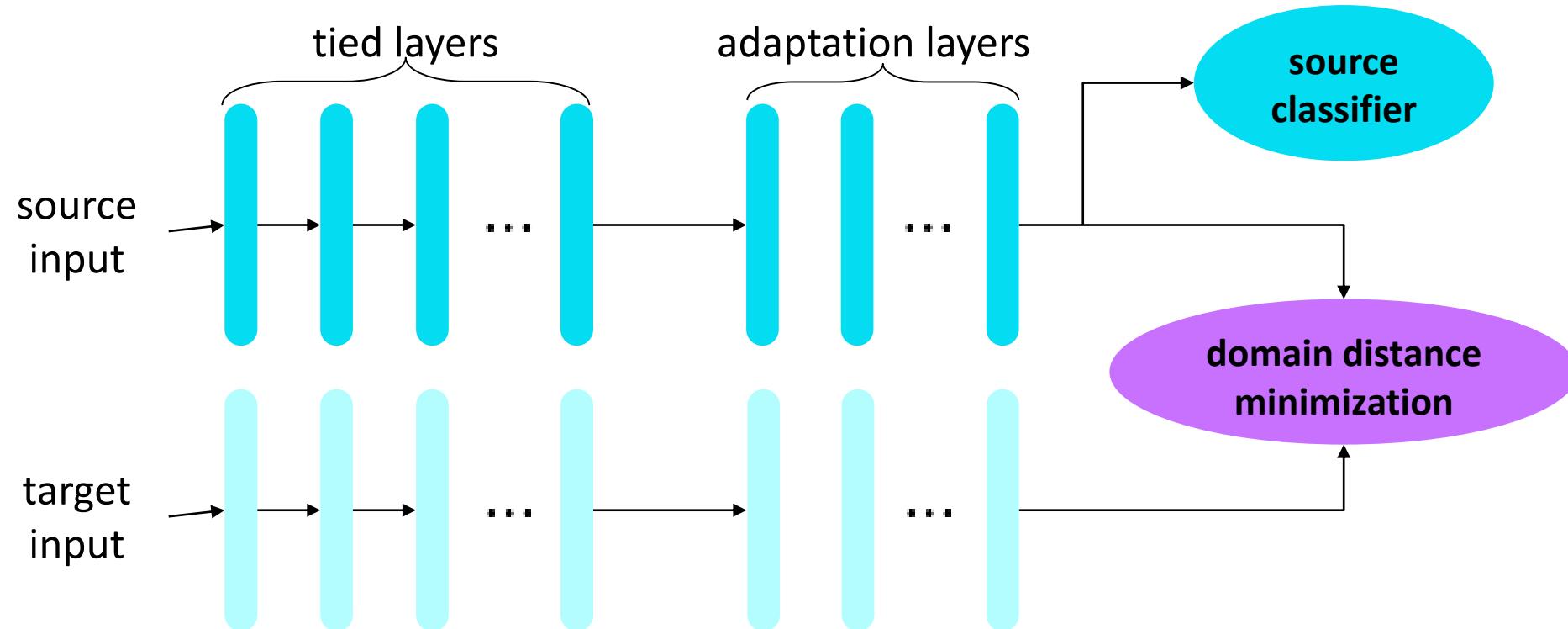
Driving in Mainland China



Driving in Hong Kong SAR, China

# Transfer Learning in a Deep Model

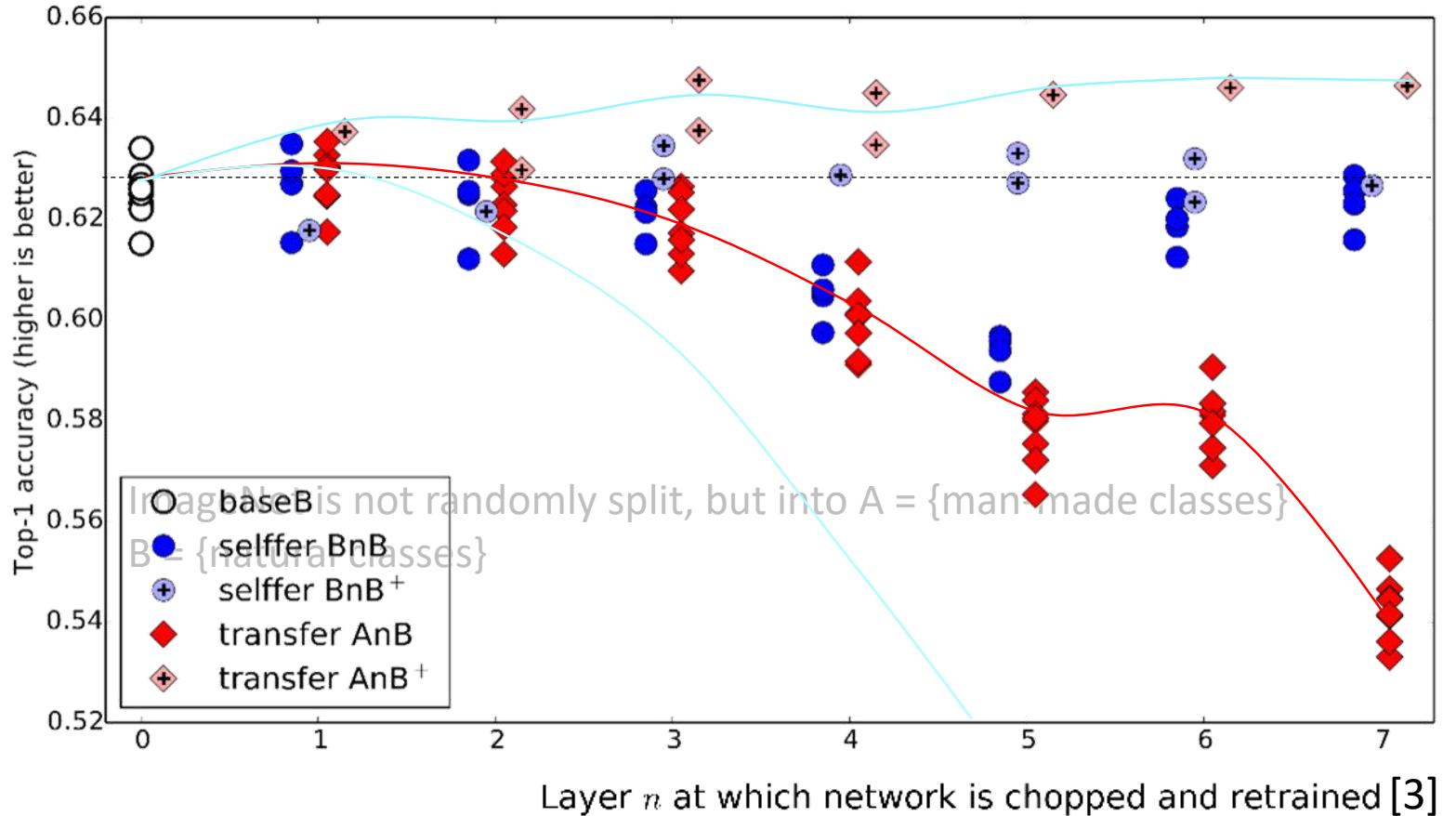
- **Objective**  $\mathcal{L} = \mathcal{L}_{\text{source}} + \mathcal{L}_{\text{distance}}$



Learning transferable features with deep adaptation networks. M Long, Y Cao, J Wang, MI Jordan. International Conference on Machine Learning (ICML) 2015

# Transfer Learning in a Deep Model

## A Quantitative Study



Conclusion: lower layer features are more general and transferrable, and higher layer features are more specific and non-transferrable.



## Transfer Learning Setting I :

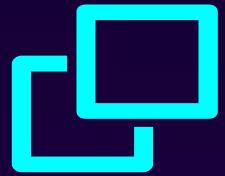
- Source domain: sufficient labeled data
- Target domain: no labeled data
- Domain Adaptation



## Transfer Learning Setting II :

- Source domain: sufficient labeled data
- Target domain: little labeled data
- Supervised Transfer Learning





# Transfer Learning Setting I

Source domain: sufficient labeled data

Target domain: no labeled data

# Sentiment Analysis

rating

★ 10/10



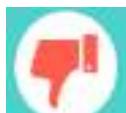
This movie will blow your mind and break your heart - and make you desperate to go back for more. Brave, brilliant and better than it has any right to be.

shawneofthedead 25 April 2018

Over the past decade, Marvel has earned itself the benefit of the doubt. The studio has consistently delivered smart, funny, brave films that both embrace and transcend their comic-book origins. The 18 blockbuster movies produced since Iron Man first blasted off into the stratosphere in 2008 have not only reinvented superhero films as a genre - they've helped to legitimise it. Indeed, Marvel's two most recent films - Thor: Ragnarok and Black Panther - have received the kind of accolades usually reserved for edgy arthouse flicks.

rating

★ 1/10



I actually laughed out loud at the end

tenaciouspeas 23 May 2018

What a trash heap of a movie. I thought about giving it 2 stars because there were a couple of things that made me chuckle but I left the theater so irritated that I talked myself out of it. I kept singing the "I don't care" song for the last 2 hours of this movie, which seemed to last at least 5 hours long. I'm sure they could have fit at least 2 more bad CGI action fight scenes in there, to make it 6 hours long. I loved the first Avengers. I loved Thor Ragnarok. I hated this movie, which can easily be summed up: A really long movie about a boring CGI character titled: Here's Thanos!

44 out of 80 found this helpful. Was this review helpful? Yes No | Report this

## ➤ Single-Domain Solution

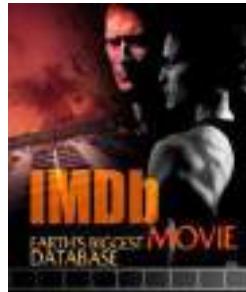
depends on sufficient labeled data

## ➤ Cross-domain solution: Transfer Learning

Transferring sentiment classification knowledge from one domain to another

# Cross-Domain Features: Pivots

Source domain (**Movie**)



Target domain (**Electronics**)



	Great movie. His characters are <b>engaging</b> and <b>thoughtful</b> .	This great touchpad feels <b>glossy</b> and is <b>responsive</b> .
	It's a <b>excellent</b> , <b>sobering</b> drama.	It is very <b>lightweight</b> , <b>excellent</b> transition from PC.
	An <b>terrible</b> movie. It is very <b>plotless</b> and <b>insipid</b> .	It is <b>blurry</b> and <b>fuzzy</b> in very dark setting. So <b>terrible</b> HP.



Domain adaptation with structural correspondence learning, Blitzer et al. EMNLP 2006

# Structural Correspondence Learning (SCL)

## ➤ Unlabeled step: **pivot predictors**

- **pseudo-label:** select M pivot features from keywords
- Each pivot predictor aligns non-pivot features from **source** to **target** domains.

## ➤ Example:

Movie

**review1:** Very **great** movie. His characters are **engaging** and **thoughtful**.

Electronics

**review2:** This **great** touchpad feels **glossy** and is **responsive**.

**Binary problem:** Does the pivot “**great**” appear in the review?

## ➤ Transformed samples:

N non-pivot features

	engaging	thoughtful	responsive	glossy	...
review1:	1	1	0	0	...
review2:	0	0	1	1	...

M pseudo labels

	great	awful	...
review1:	1	0	...
review2:	1	0	...

**Movie**

	engaging thoughtful responsive glossy	Sobering lightweight	Plotless insipid Blurry fuzzy
1	1	0	0
0	0	1	0
-1	0	0	1



**Training**

$$y=f(\mathbf{x}) = \text{sgn}(\mathbf{w}\mathbf{x}^T), \mathbf{w} = [1, 1, -1]$$

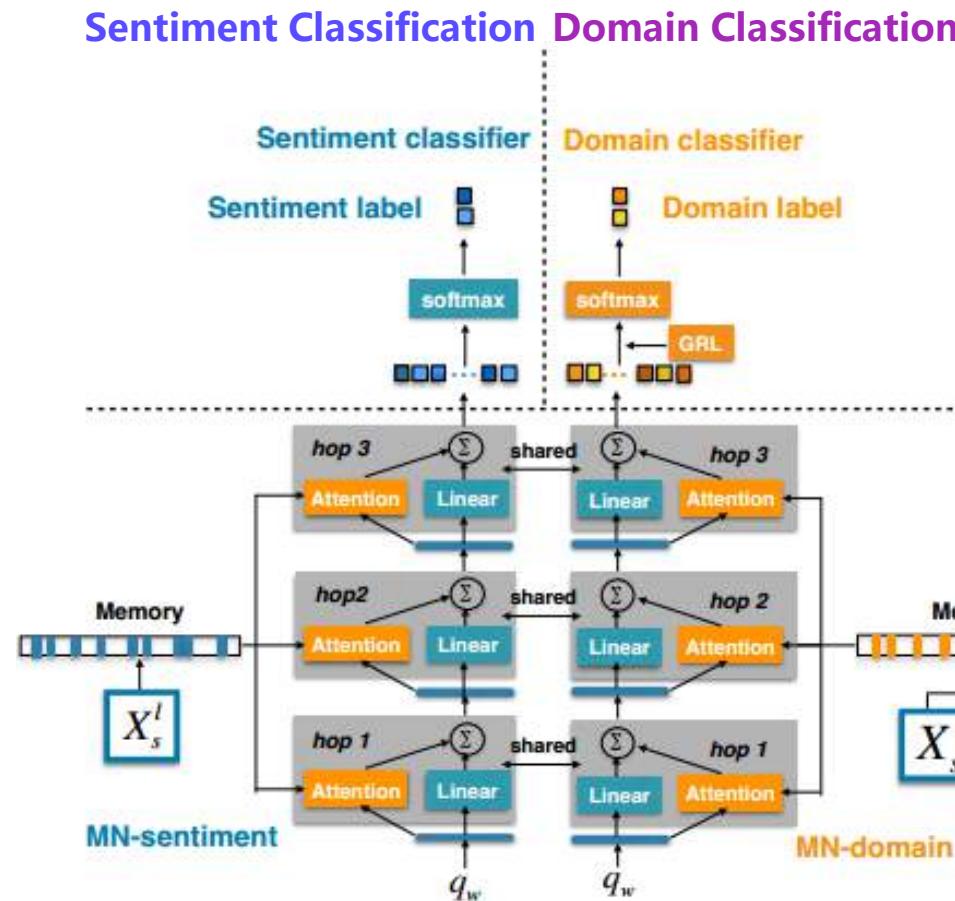


**Prediction**

**Electronics**

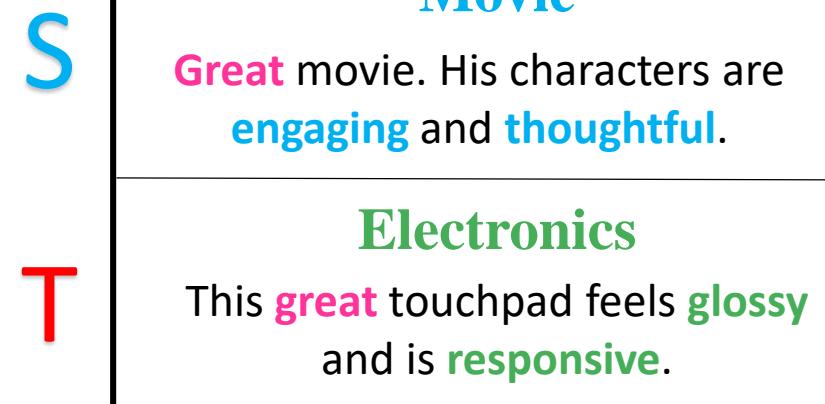
	engaging thoughtful responsive glossy	Sobering lightweight	Plotless insipid Blurry fuzzy
1	1	0	0
0	0	1	0
-1	0	0	1

# An Adversarial Approach



**Domain Classification Objective:**  
Maximize domain classification error

Source data  $X_s$   
Target data  $X_t$



Li, Zheng, Qiang Yang, et al. "End-to-end adversarial memory network for cross-domain sentiment classification." IJCAI 2017.

# Comparison with baseline methods

## ✓ Traditional methods:

**SCL**: Structural Correspondence Learning [Blitzer et al., 2006]

**SFA**: Spectral Feature Alignment [Pan et al., 2010]

## ✓ AMN model significantly outperforms the traditional methods SFA and SCL on Amazon Reviews Dataset

GT1 Prediction1  
great gift i have bought over 100 of these in the past 6 months i have only had 1 burn badly haven't found a dslr player yet that they won't play in

GT1 Prediction1  
good for canon a85 ~~camera~~ take all the videos and pictures you want with the best quality

GT1 Prediction1  
you cannot beat a belkin cable great quality ~~wooden~~ construction and strong rj45 plugs i have worked with a decent share of cat6 and i have never had to cut and terminate a belkin cable due to regular wear and tear

GT1 Prediction0  
i can hear your sound output is terrible you can't hear it in a car or airplane with high quality noise canceling headphones when i called customer service they told me it was not intended for use in a car or airplane picture is very good but i have heard better sound from much cheaper players don't waste your money

GT1 Prediction0  
great technology ~~initial~~ customer experience i had the same exact experience with the ~~poor~~ fit of these headphones and the really customer service their surround sound is 99% phones don't fit well either

GT1 Prediction0  
~~not~~ terrible i had these headphones for a few years then they got crushed in half in my bag they last your ears after about ten minutes they are durable though i would recommend the kind that clip behind your ear

(a) Electronics domain

GT1 Prediction1  
great gift i love the rapid ice wine coolers i give them for token gifts and use them frequently myself they are great for a spur of the moment glass of wine that needs chilling

GT1 Prediction1  
an elegant way of serving its a traditional serve ware for serving the soup rotates the color of the turns set allows it to be used with many of the dinersware atop the side is adequate to serve at least 8-10 people the under plate is something not found with usual turner sets which gives it an elegant look but it appears a little overpriced

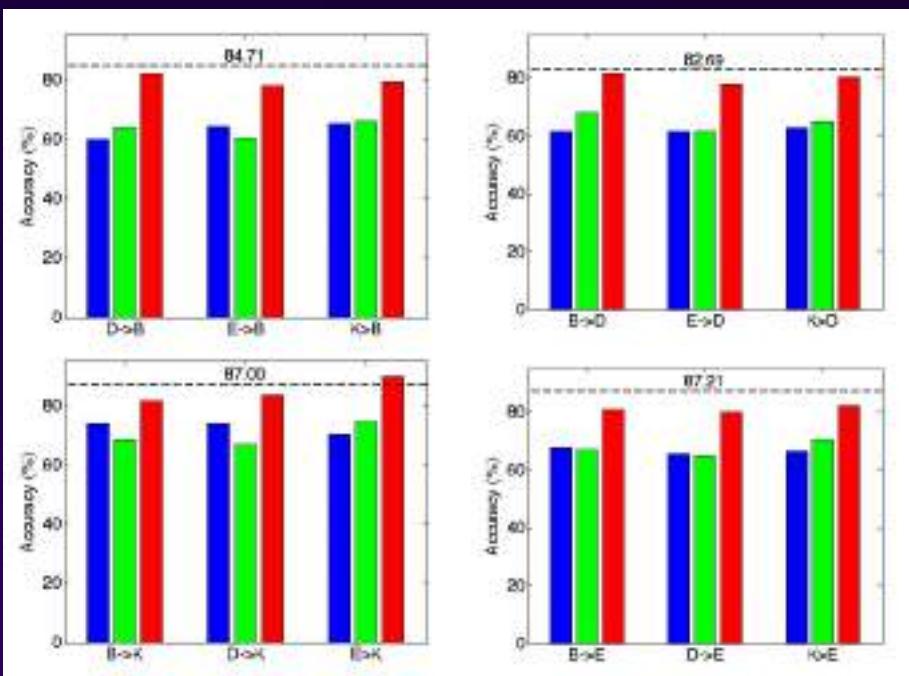
GT1 Prediction1  
2010 gift i just received this as a wedding gift and it is beautiful a great gift

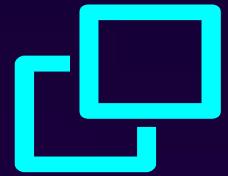
GT1 Prediction0  
~~disappointed~~ - worker i am usually very pleased with eco products but this one is a big disappointment i have not found it to be good for at anything wished it served the five backs

GT1 Prediction0  
too ~~poorly~~ made for everyday use we have a full line of flatware dishware and thought having the matching flatware would be nice after a year of standard use and discovering about 13 of the flatware is ~~missing~~ the upside is that it is cheap and replaceable but count me among those who would rather pay more for something that lasts we see in the process of ditching the flatware line and moving to something more robust

GT1 Prediction0  
totally ~~satisfied~~ we bought this to use at events for a chocolate themed group at college and used it several times before giving up

(b) Kitchen domain



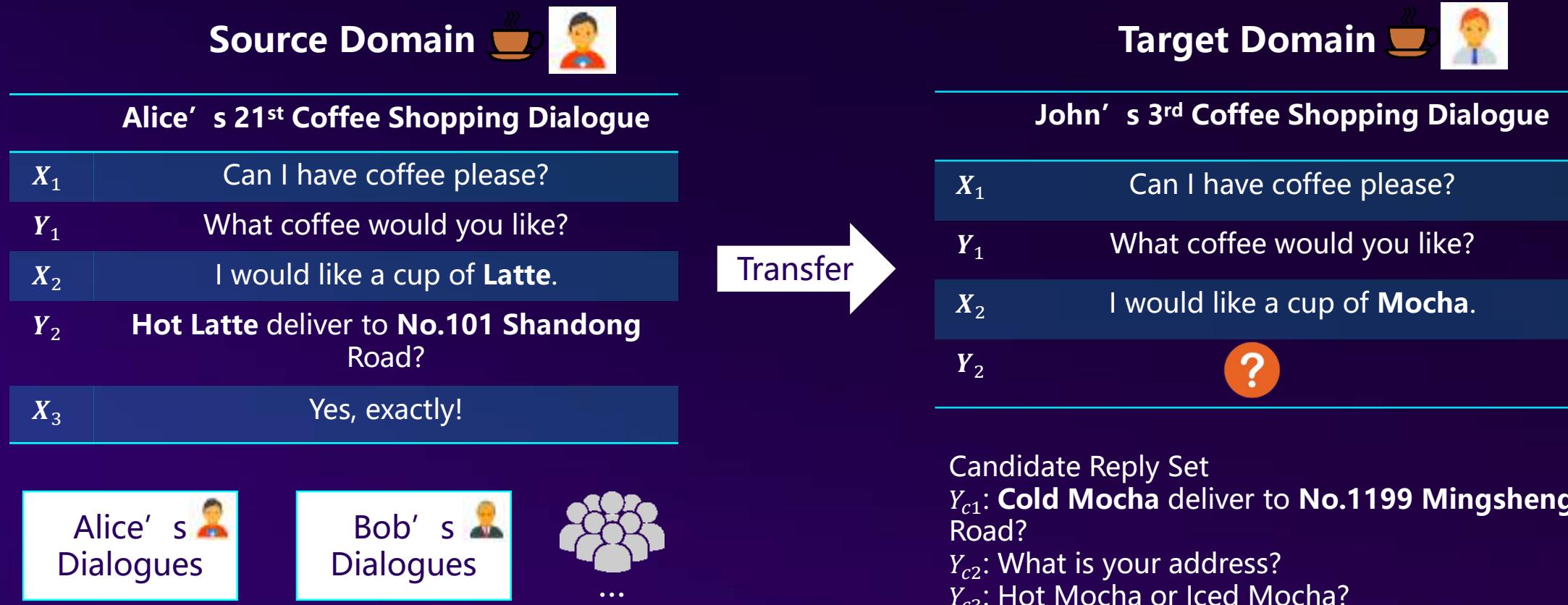


## Transfer Learning Setting II : Supervised Transfer Learning

Source domain: sufficient labeled data

Target domain: little labeled data

# Transfer Learning in Dialog Systems



# Learning Common Dialogue States and a Personalized Q-function

- Common dialogue states are learned in a source domain

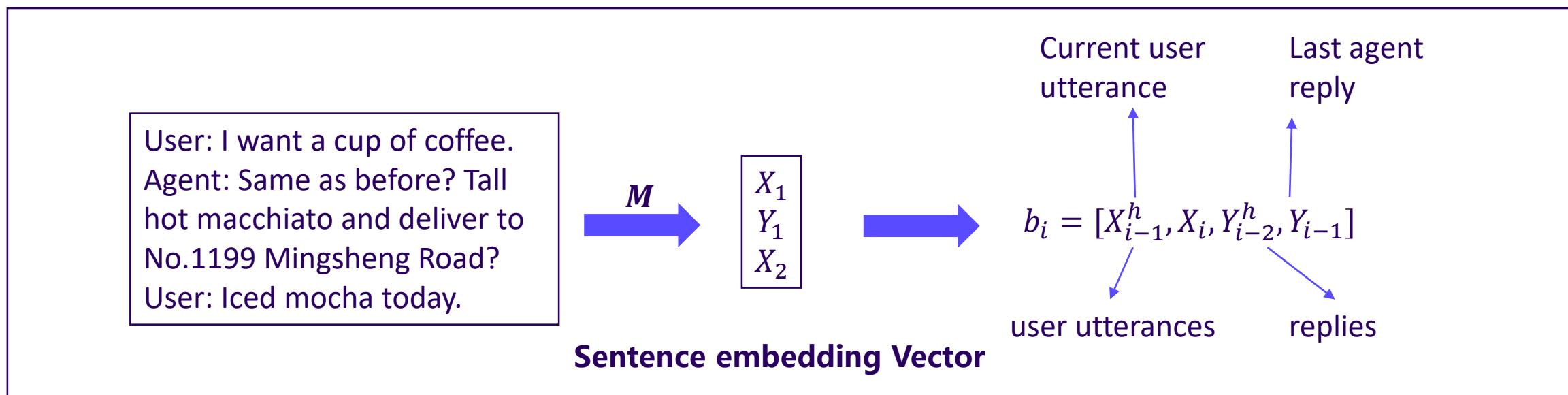
Belief state vector  $b_i = f(H_i; \mathbf{M})$ , where dialogue history  $H_i = \{\{X_j, Y_j\}_{j=1}^{i-1}, X_i\}$

- Personalized Q-function

$$Q^{\pi_u}(H_i, Y_i | \Theta) = Q_g(H_i, Y_i | \Theta^g) + Q_p(H_i, Y_i | \Theta_u^p).$$

General part      Personal part

General part:  $Q_g(H_i, Y_i | \Theta^g)$ , personal part:  $Q_p(H_i, Y_i | \Theta_u^p)$



# Dialogue Policy Transfer Examples

- Transfer across users

 	$X_1$ Can I have coffee please?
	$Y_1$ Same as before? Medium Hot Latte and deliver to No.101 Shandong Road?
	$X_2$ Yes, as always. <Success-Feedback>

State=[ $X_1$ ]  $\longrightarrow$  State=[ $X_1, Y_1, X_2$ ]  
Action= $Y_1$  Action= $Y_2$  <Success>



 	$X_1$ Can I have coffee please?
	$Y_1$ Same as before? Tall Iced Mocha and deliver to No.1199 Mingsheng Road?

State=[ $X_1$ ]  
Action= $Y_1$

- Transfer across domains



 	$X_1$ I am looking for a restaurant.
	$Y_1$ What food would you like?

State=[Food=? , Area=? , Price=?]  
Action=Request(Food)



 	$X_1$ I want to find a hotel.
	$Y_1$ How many stars should the hotel have?

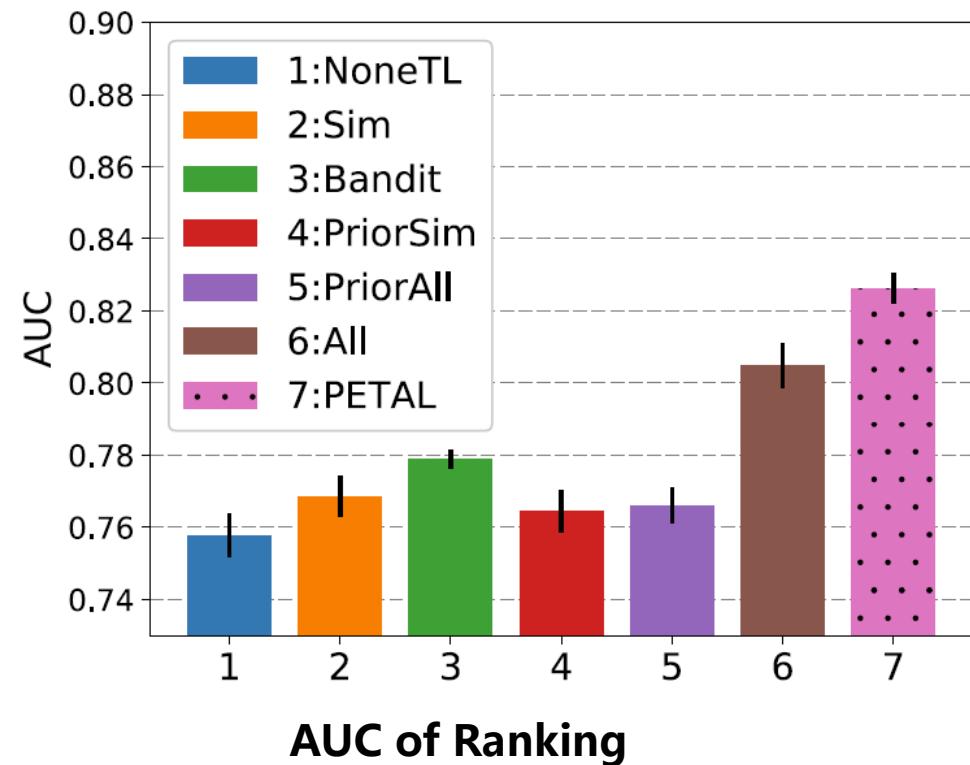
State=[Type=? , Stars=? , Area=? , Park=?]  
Action=Request(Stars)

# Real-world Experiment

## ➤ Setting: Coffee ordering

- Collected in O2O company, between real customers and human personal assistants.
- 52 source users and 20 target users, 2000+ multi-turn dialogues.
- Evaluation: AUC

	Source Domain		Target Domain	
	Users	Dialogues	Users	Dialogues
Real Data	52	1859	20	329
Simulation	11	176000	5	100



User utterance :	I want a cup of coffee.	
All	PETAL	Response Candidates
0.86	1.36	* Same as before? Tall hot americano and deliver to Central Conservatory of Music?
0.99	0.92	All right, deliver to No.1199 Beiyuan Road, Chaoyang District, Beijing?
0.72	0.69	What's your address?

## Recommendation System

### Supervised learning based RecSys

Single Domain RecSys easily get stuck in local optimal and keeps recommending the similar

- Performs poor for new user, new article, and new domain.
- Insensitive to fast evolving user interests, explores leading to worse short-term CTR.

### Transferable Contextual Bandit

王者荣耀里五个单体伤害最高的英雄，他曾经一个技能秒掉后羿！

天云解说



刚刚，苹果扔出重磅炸弹，华为或要哭了！

老板思维首府



终于看懂了，《天龙八部》原来就是一场杀人游戏

好书推荐

六神磊磊读金庸



王者荣耀里五个单体伤害最高的英雄，他曾经一个技能秒掉后羿！

天云解说



王者荣耀里五个单体伤害最高的英雄，他曾经一个技能秒掉后羿！

天云解说



王者荣耀里五个单体伤害最高的英雄，他曾经一个技能秒掉后羿！

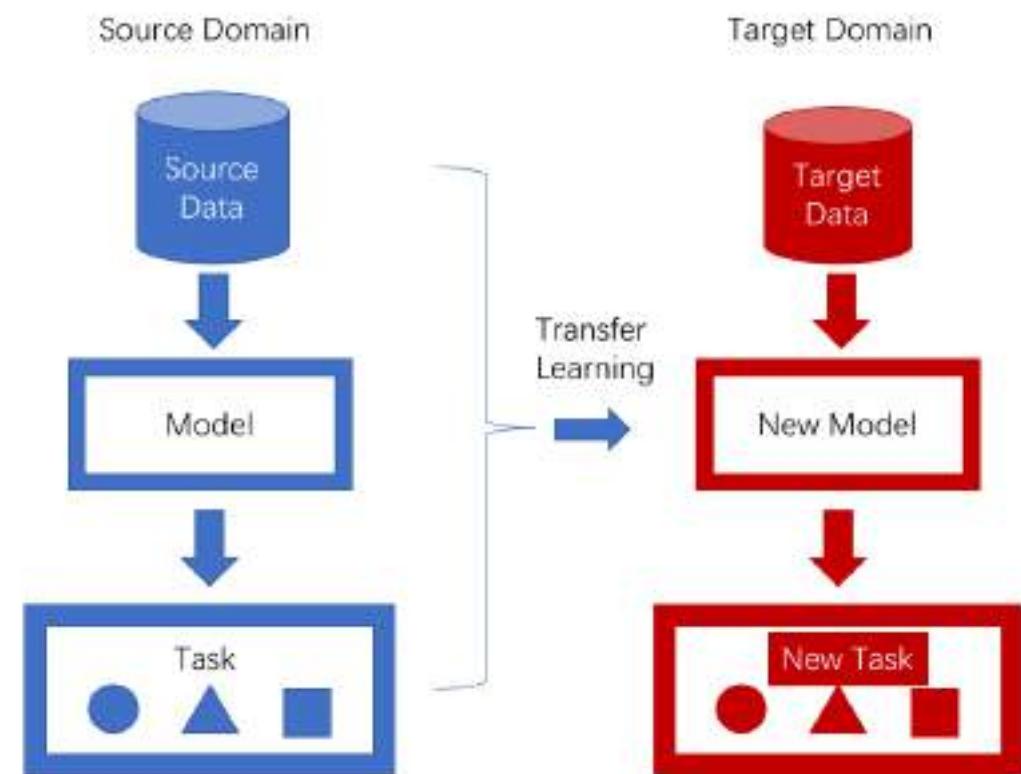
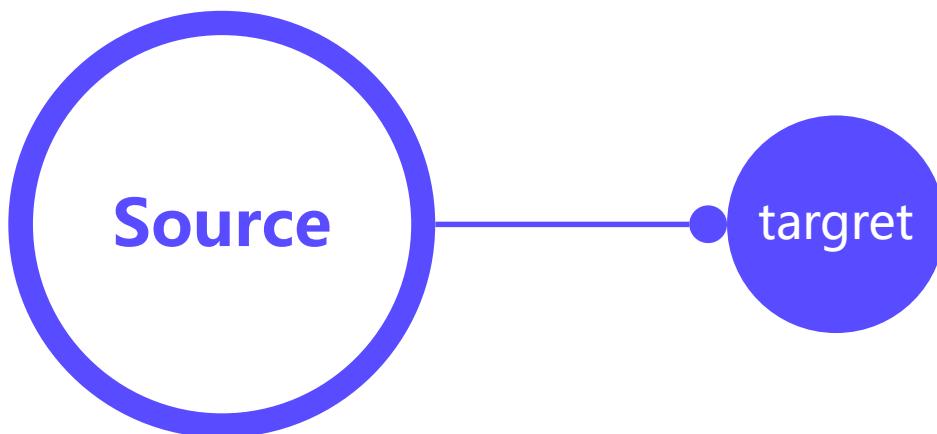
天云解说



Transferable Contextual Bandit for Cross-Domain Recommendation, Bo Liu, Yu Zhang, Qiang Yang et al.  
AAAI18

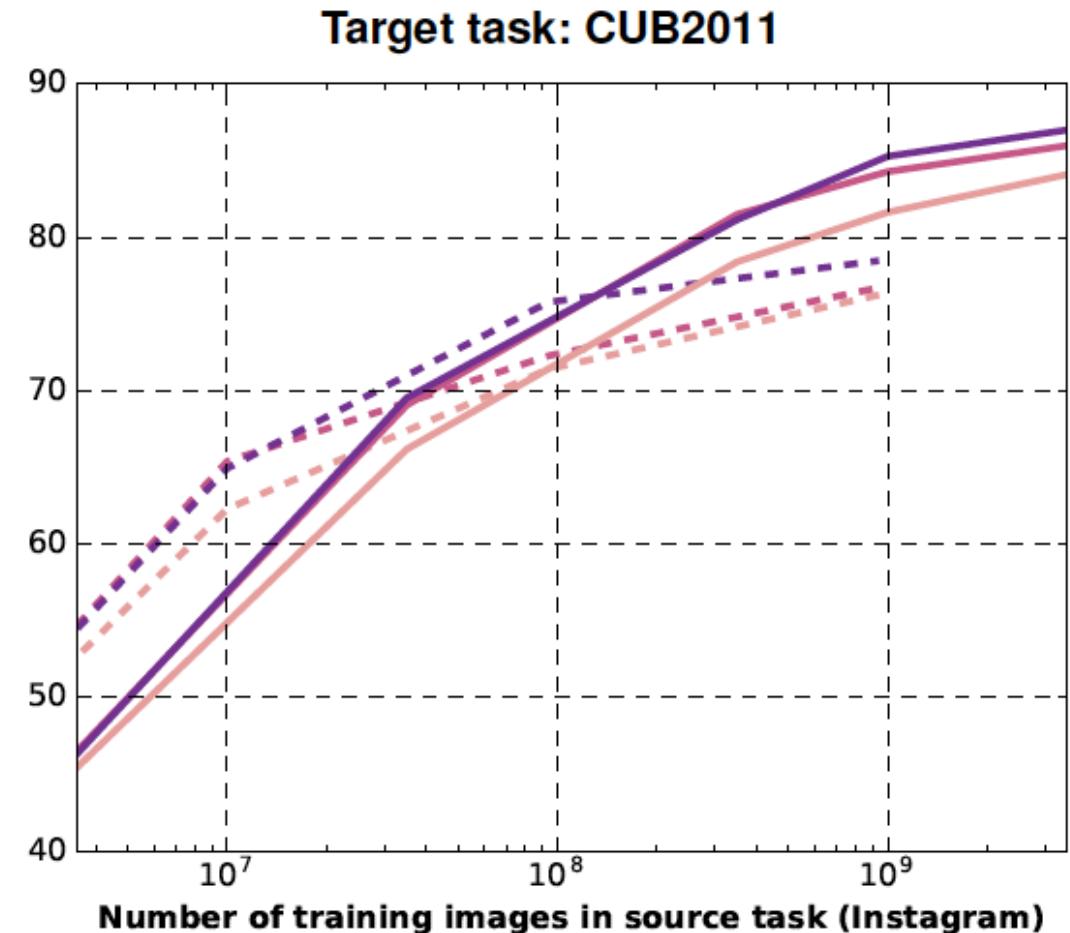
# Trend in Transfer Learning : Using Huge Pretrained Model

- Source domain: **huge** labeled or unlabeled data
- Target domain: few labeled data
- Objective: transfer model from source domain to target domain **for same or different tasks**



# Source-Data Scale Matters in Transfer Learning (image)

- Dhruv Mahajan, et al.: **Exploring the Limits of Weakly Supervised Pretraining**. ECCV (2) 2018
- "Without manual dataset curation or sophisticated data cleaning, models trained on billions of Instagram images using thousands of distinct hashtags as labels exhibit excellent transfer learning performance"



# Scale of Source-Data Matters in Transfer Learning (NLP): BERT

Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina

Toutanova: BERT: Pre-training of Deep Bidirectional  
Transformers for Language Understanding. CoRR  
abs/1810.04805 (2018)

“ Recent empirical improvements due to transfer learning with language models have demonstrated that rich, unsupervised pre-training is an integral part of many language understanding systems.

Our major contribution is further generalizing these findings to deep bidirectional architectures, allowing the same pre-trained model to successfully tackle a broad set of NLP tasks.”

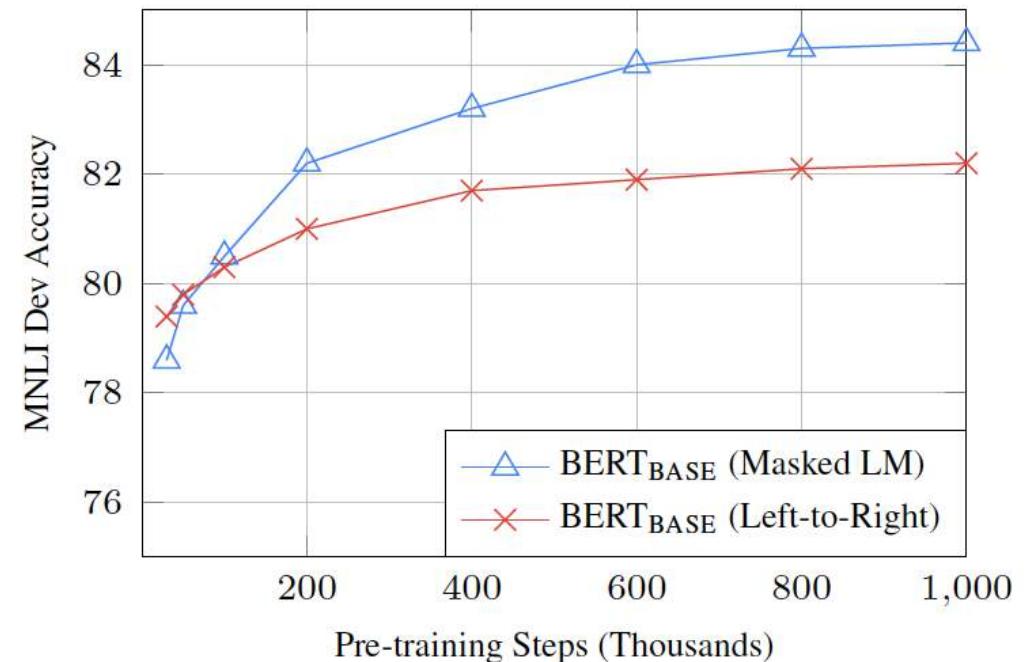
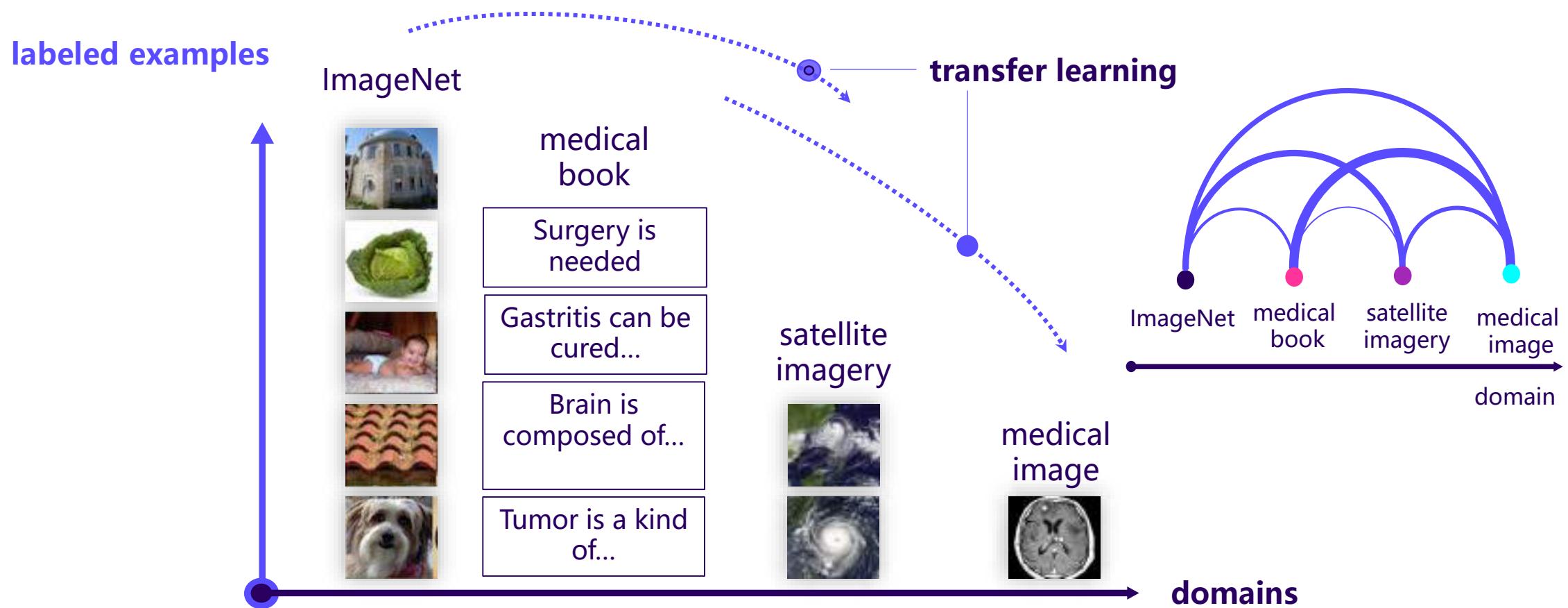


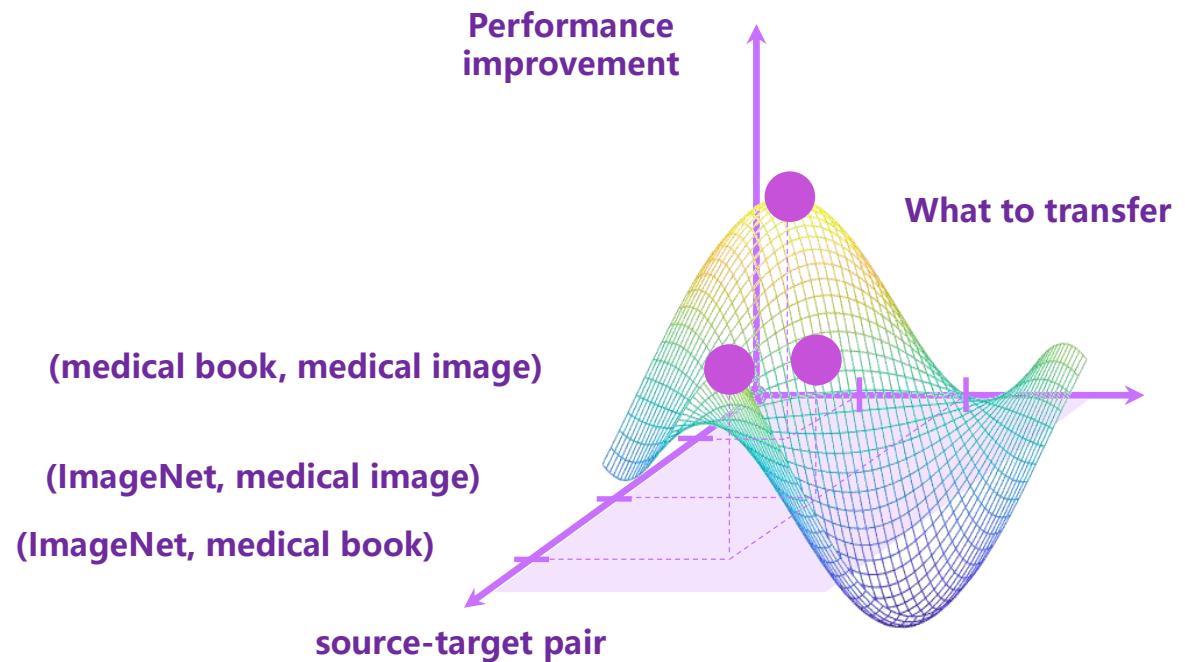
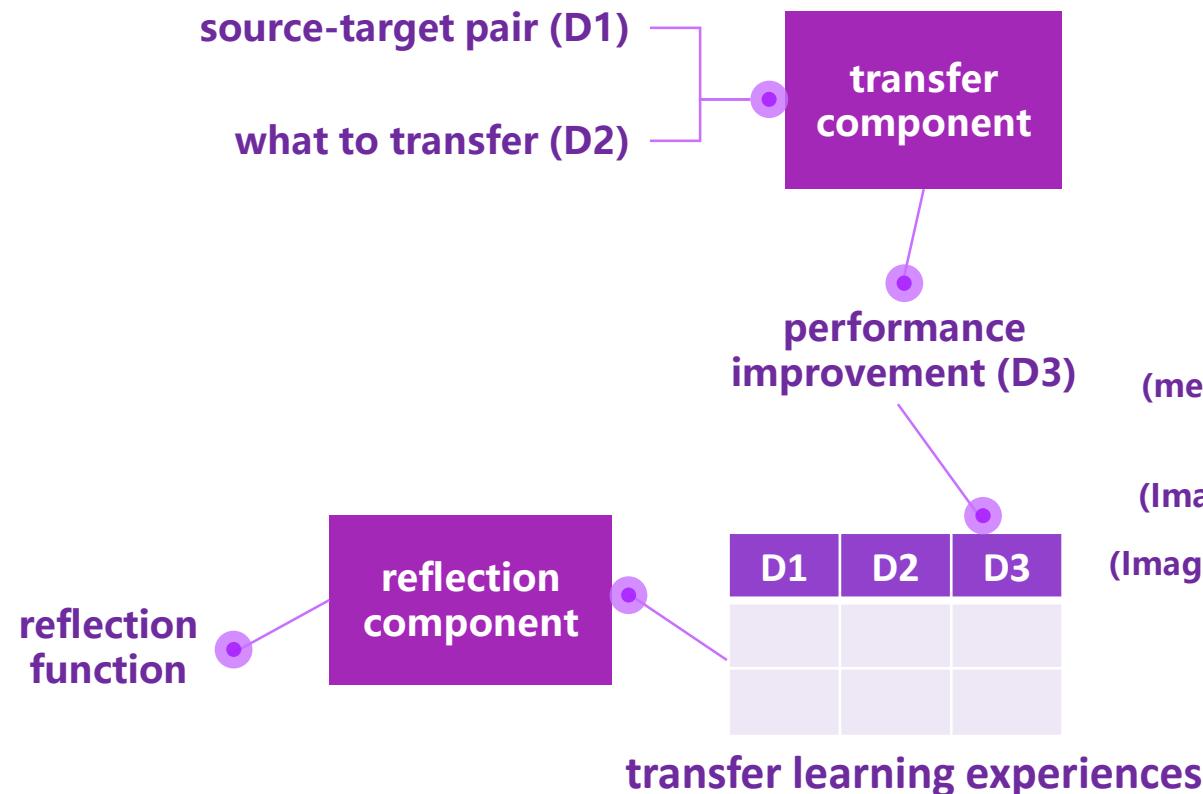
Figure 4: Ablation over number of training steps. This shows the MNLI accuracy after fine-tuning, starting from model parameters that have been pre-trained for  $k$  steps. The x-axis is the value of  $k$ .

# Transfer Learning via Learning to Transfer

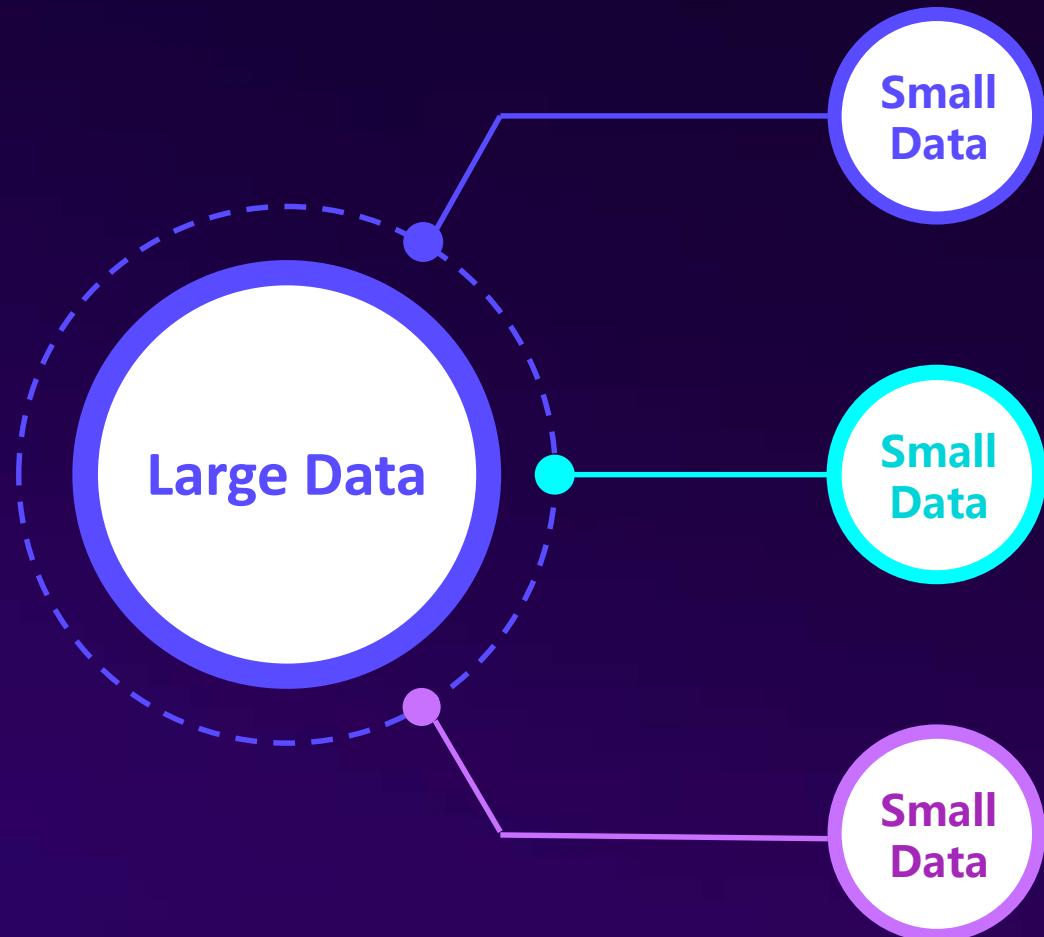


# Learning-to-Transfer (L2T) Framework

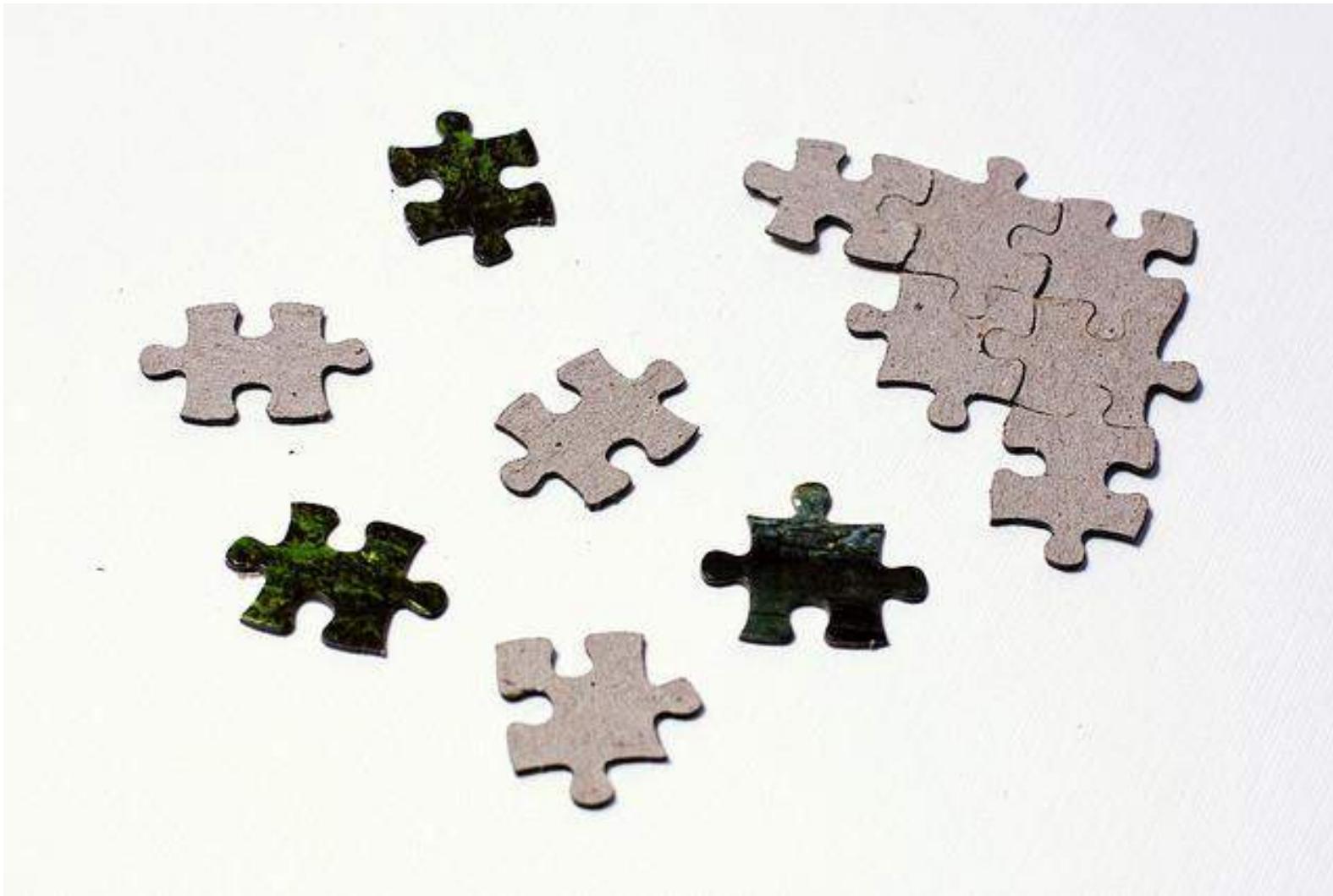
## ➤ Training – Learning skills from experiences



# Transfer Learning from Large Data to Small Data



## Next Problem: Data Are Fragmented

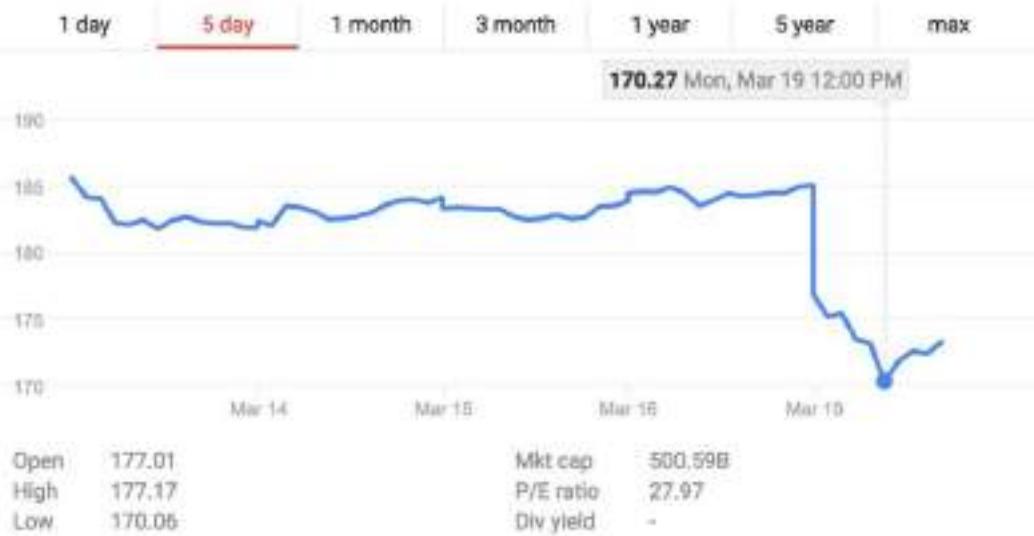


# Challenges to AI: Data Privacy and Confidentiality

## *Facebook's data privacy scandal*

Market summary > Facebook, Inc. Common Stock  
NASDAQ: FB - Mar 19, 2:21 PM EDT

172.32 USD **-\$12.77 (6.90%)**



**2019/1/19**

FTC reportedly planning 'record-setting' fine against Facebook for mishandling user data

Charles Miller · Jan 20th 2019 7:31 am ET · @CharlesMiller



- More than 50 million people involved
- UK assessed a £500,000 fine to Facebook
- **the worst single-day market value decrease for a public company in the US, dropping \$120 billion, or 19%**
- In 2012, the FTC fined Google \$22.5 million over failing to improve privacy practices – a record for such a punishment.
- The Washington Post says that the fine against Facebook is expected to be “much larger.”

# The General Data Protection Regulation (GDPR)



- No Autonomous Modeling and Decision
- Interpretability of Model Decisions
- Users'Right for Data to be Forgotten
- Data Privacy By Design
- Explicit Consent for Data Usage

# California Consumer Privacy Act (CCPA)

- Takes effect in 2020
- grants consumers the right to know what information is collected and **whom it is shared with**
- Consumers will have the option of barring tech companies from selling their data
- Provides some of the strongest **regulations in the USA.**



# China's Data Cyber Security Law

- Enacted in 2017
- Requires that Internet businesses must not leak or tamper with the personal information
- When conducting data transactions with third parties, they need to ensure that the proposed contract follow legal data protection obligations.
- More to come...

## Highlights and interpretation of the Cybersecurity Law



### Highlights of the Cybersecurity Law

Comprising 79 articles in seven chapters, the Cybersecurity Law contains a number of cybersecurity requirements, including safeguards for national cyberspace sovereignty, protection of critical information infrastructure and data and protection of individual privacy. The Law also specifies the cybersecurity obligations for all parties. Enterprises and related organisations should prioritise the following highlights of the Cybersecurity Law:



#### Personal Information protection

The Cybersecurity Law clearly states requirements for the collection, use and protection of personal information.



#### Critical information infrastructure

The Cybersecurity Law frequently mentions the protection of 'critical information infrastructure'.



#### Network operators

'Network operators' are the owners and administrators of networks and network service providers. The Cybersecurity Law clarifies operators' security responsibilities.



#### Preservation of sensitive information

The Cybersecurity Law requires personal information/important data collected or generated in China to be stored domestically.



#### Certification of security products

Critical cyber equipment and special cybersecurity products can only be sold or provided after receiving security certifications.

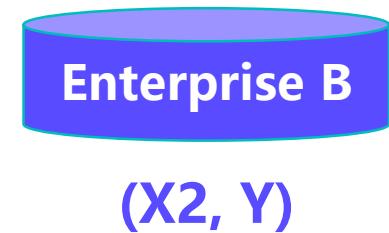
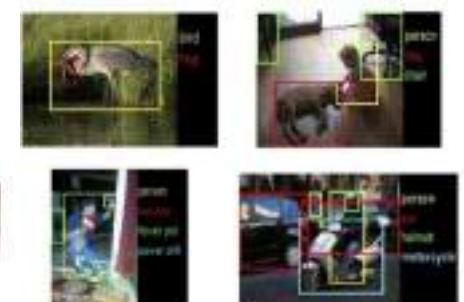
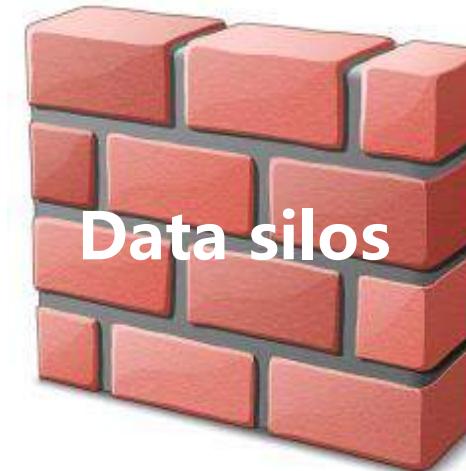
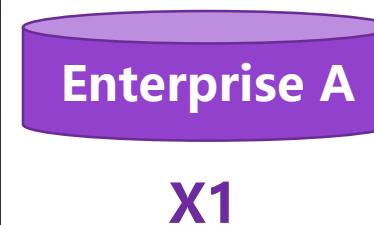


#### Legal liabilities

Enterprises and organisations that violate the Cybersecurity Law may be fined up to RMB1,000,000.

From Report by KPMG 2017

# Challenges to AI : small data and fragmented data



Low Security in Data Sharing  
Lack of Labeled Data  
Segregated Datasets

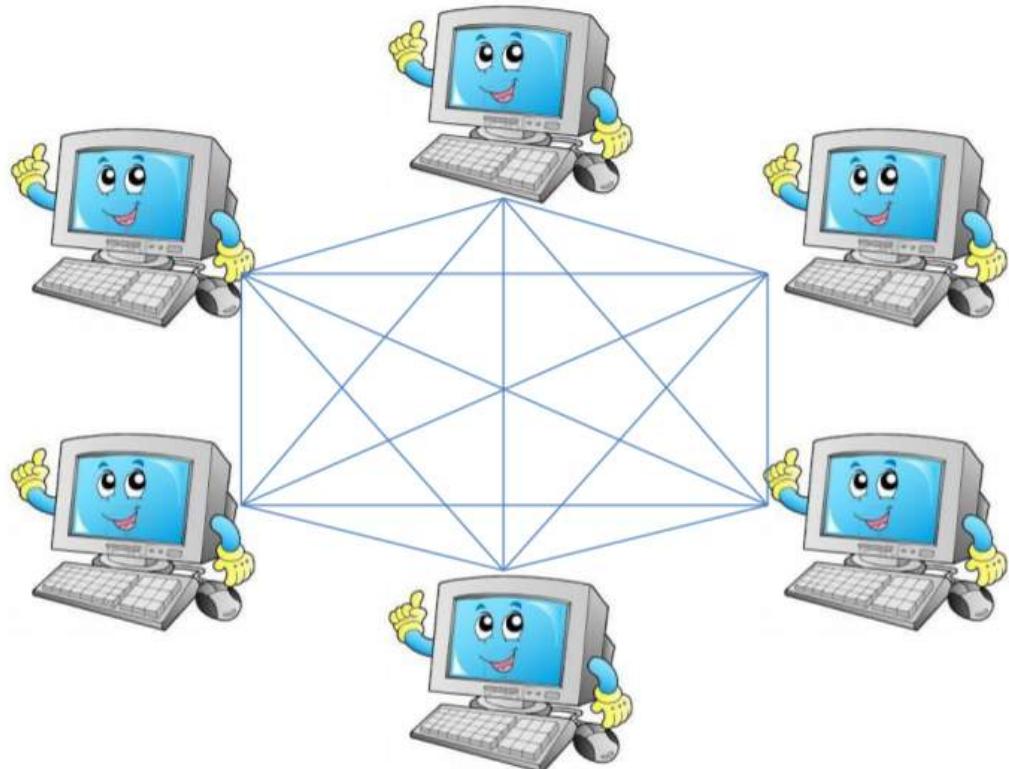
**Over 80% of enterprises' information in data silos!**

# Privacy-Preserving Technologies

- **Secure Multi-party Computation (MPC)**
  - **Homomorphic Encryption (HE)**
  - **Yao' s Garbled Circuit**
  - **Secret sharing**
  - **Differential Privacy (DP)**
- .....

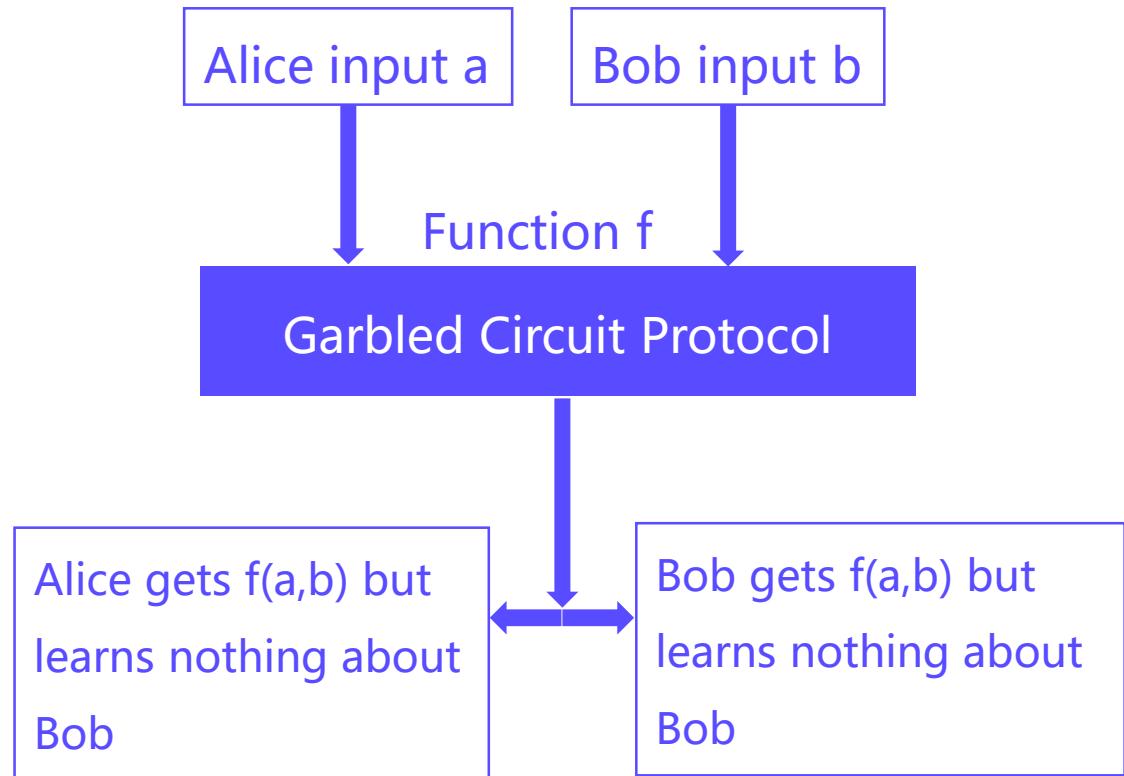


# Secure Multi-Party Computation (MPC)



- Provides security proof in a well-defined simulation framework
- Guarantees complete zero knowledge
- Requires participants' data to be secretly-shared among non-colluding servers
- Drawbacks:
  - Expensive communication,
  - Though it is possible to build a security model with MPC under lower security requirement in exchange for efficiency

# Yao' s Garbled Circuit Protocol (Andrew Yao, 1986)



- **Oblivious Transfer**

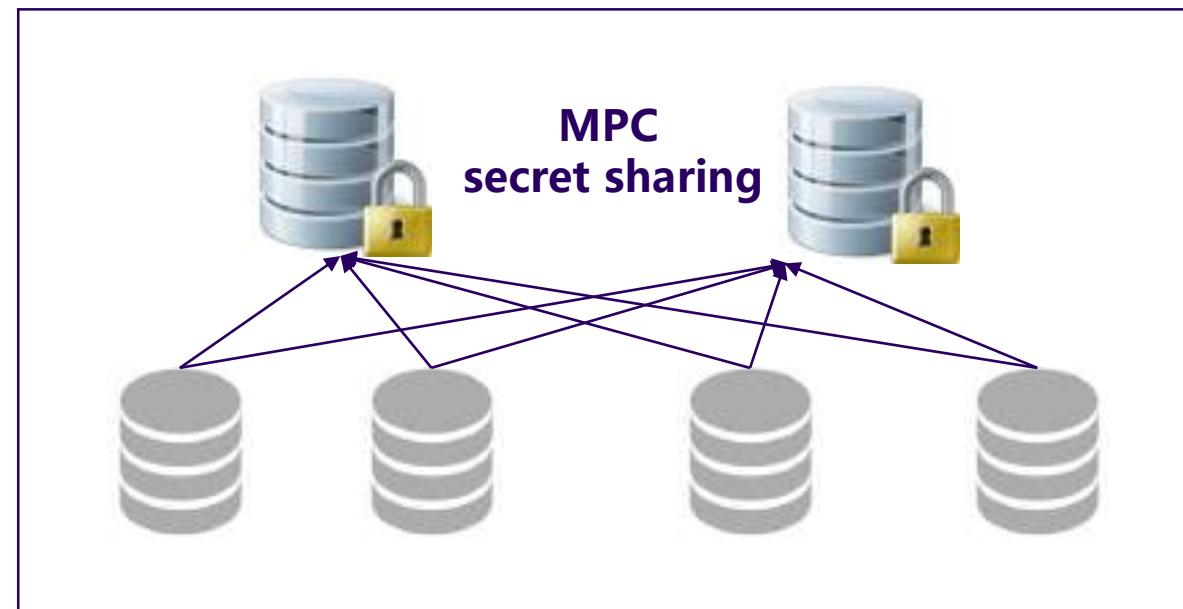


### Steps

- Alice builds a garbled circuits;
- Alice sends her input keys;
- Alice and Bob do Oblivious Transfer;
- Bob gets the output and sends back to Alice;
- Alice and Bob learns nothing about the other value.

# SecureML : Privacy-preserving machine learning for linear regression, logistic regression and neural network training

- Combines secret sharing, garbled circuits and oblivious transfer
- Learns via two un-trusted, but non-colluding servers
- Computationally expensive



**Mohassel, P., & Zhang, Y. (2017, May). SecureML: A system for scalable privacy-preserving machine learning. In *2017 38th IEEE Symposium on Security and Privacy (SP)* (pp. 19-38). IEEE.**

# Homomorphic Encryption

- Full Homomorphic Encryption and Partial Homomorphic Encryption.
- **Paillier** partially homomorphic encryption

**Addition :**  $[[u]] + [[v]] = [[u+v]]$

**Scalar multiplication:**  $n[[u]] = [[nu]]$

- For public key  $pk = n$ , the encoded form of  $m \in \{0, \dots, n - 1\}$  is

$$\text{Encode}(m) = r^n (1 + n)^m \bmod n^2$$

$r$  is randomly selected from  $\{0, \dots, n - 1\}$ .

- For float  $q = (s, e)$ , encrypt  $[[q]] = ([[s]], e)$ , here  $q = s\beta^e$  is base- $\beta$  exponential representation.

**Rivest, R. L.; Adleman, L.; and Dertouzos, M. L. 1978. On data banks and privacy homomorphisms. Foundations of Secure Computation, Academia Press 169–179.**

# Applying HE to Machine Learning

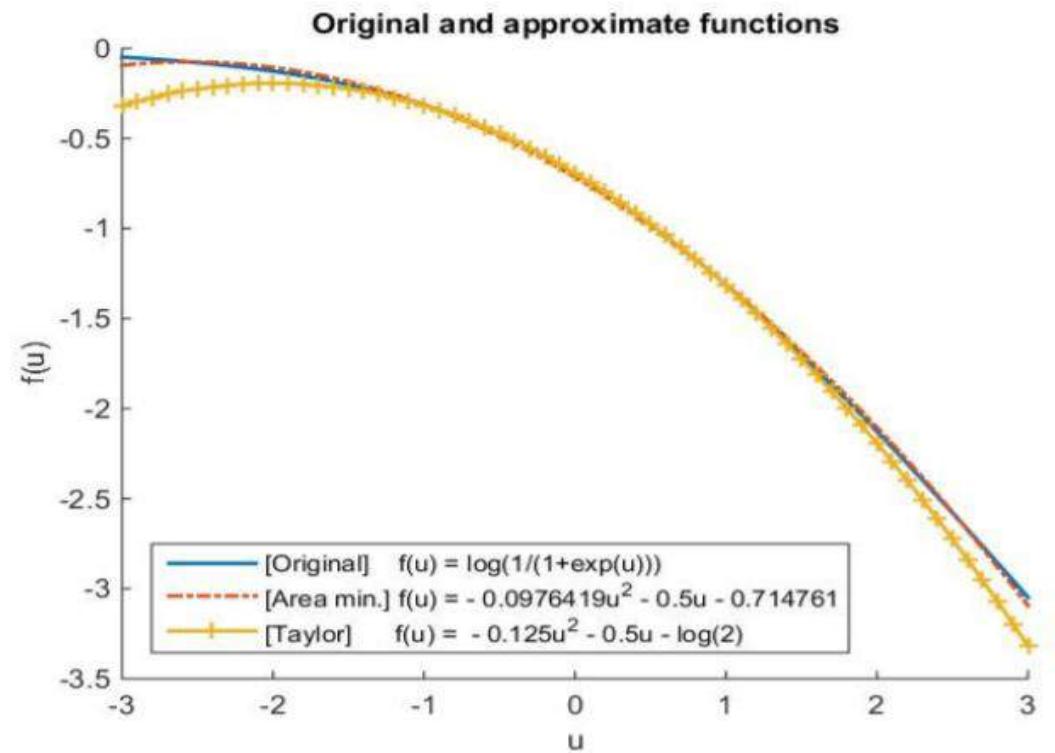
## Polynomial approximation for logarithm function

$$\log\left(\frac{1}{1 + \exp(u)}\right) \approx \sum_{j=0}^k a_j u^j$$

Encrypted computation for each term in the polynomial function

$$loss = \log 2 - \frac{1}{2} y w^T x + \frac{1}{8} (w^T x)^2$$

$$[[loss]] = [[\log 2]] + \left(-\frac{1}{2}\right) * [[yw^T x]] + \frac{1}{8} [[(w^T x)^2]]$$



- Kim, M.; Song, Y.; Wang, S.; Xia, Y.; and Jiang, X. 2018. Secure logistic regression based on homomorphic encryption: Design and evaluation. JMIR Med Inform 6(2)
- Y. Aono, T. Hayashi, T. P. Le, L. Wang, Scalable and secure logistic regression via homomorphic encryption, CODASPY16

# Is the Gradient Info Safe to Share?

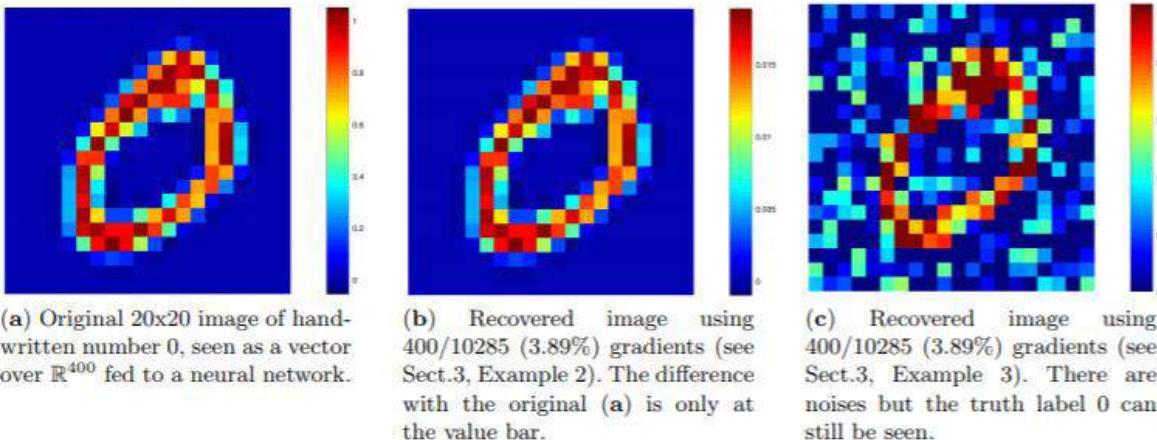
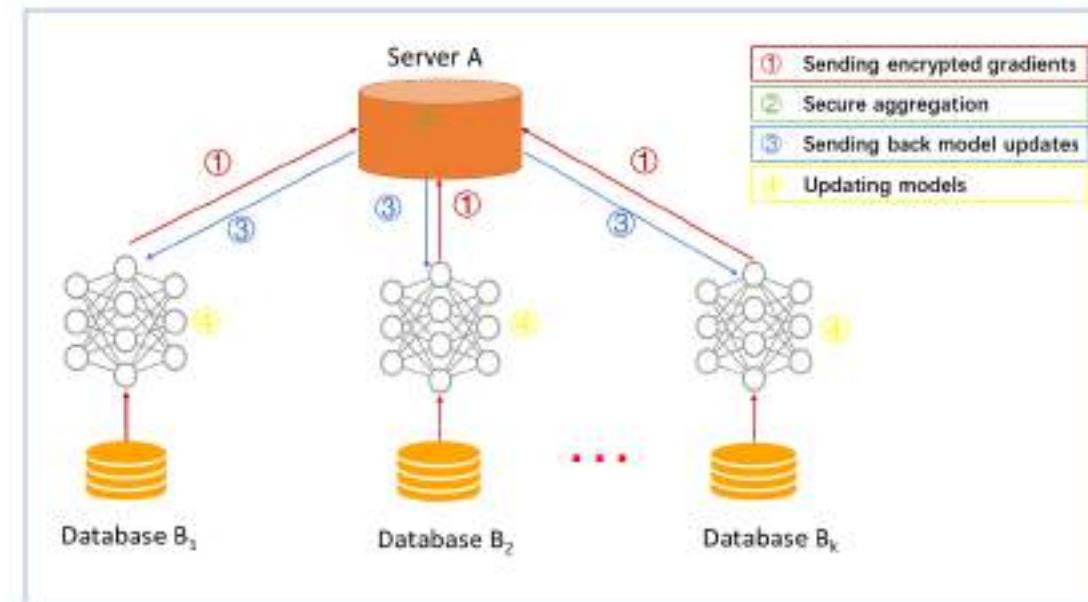


Fig. 3. Original data (a) vs. leakage information (b), (c) from a small part of gradients in a neural network.

*Protect gradients with Homomorphic Encryption*



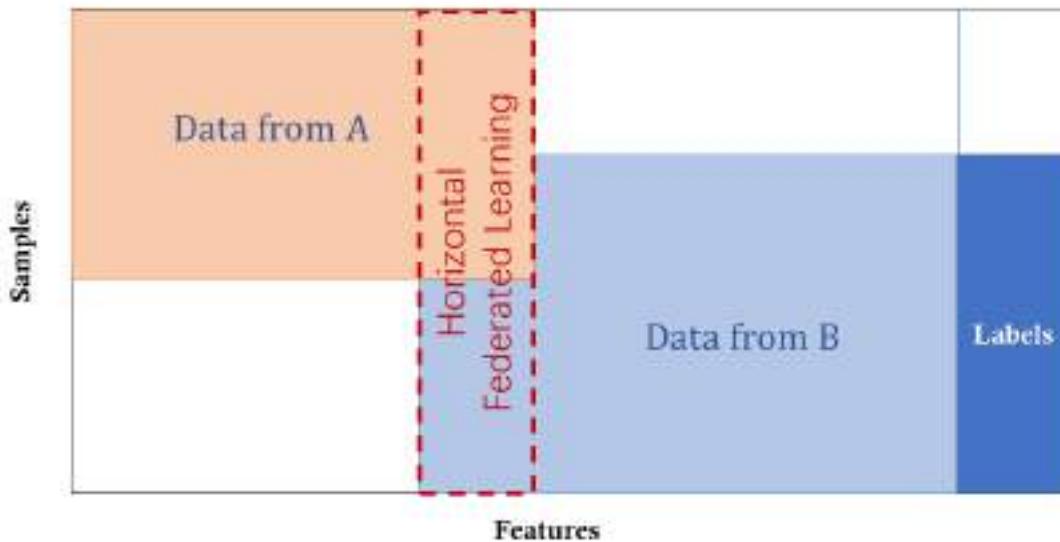
- Le Trieu Phong, et al. 2018. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. IEEE Trans. Information Forensics and Security , 13, 5 (2018), 1333–1345

- **Algorithm ensures that no information is leaked to the semi-honest server, provided that the underlying additively homomorphic encryption scheme is secure\*.**

\* Q. Yang, Y. Liu, T. Chen, Y. Tong, Federated machine learning: concepts and applications, ACM TIST, ,2018

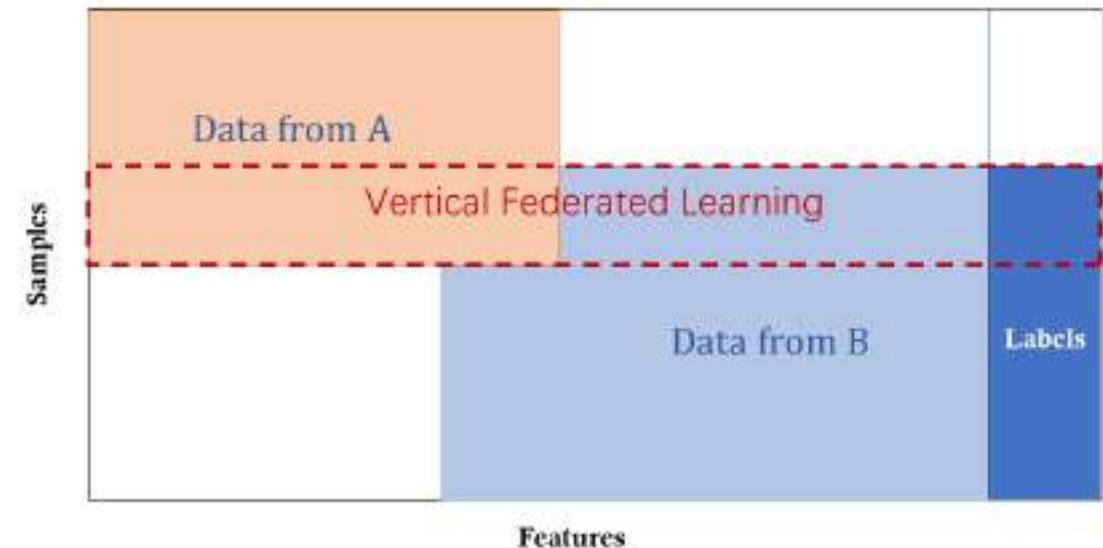
# Categorization of Federated Machine Learning

**Horizontal FML**



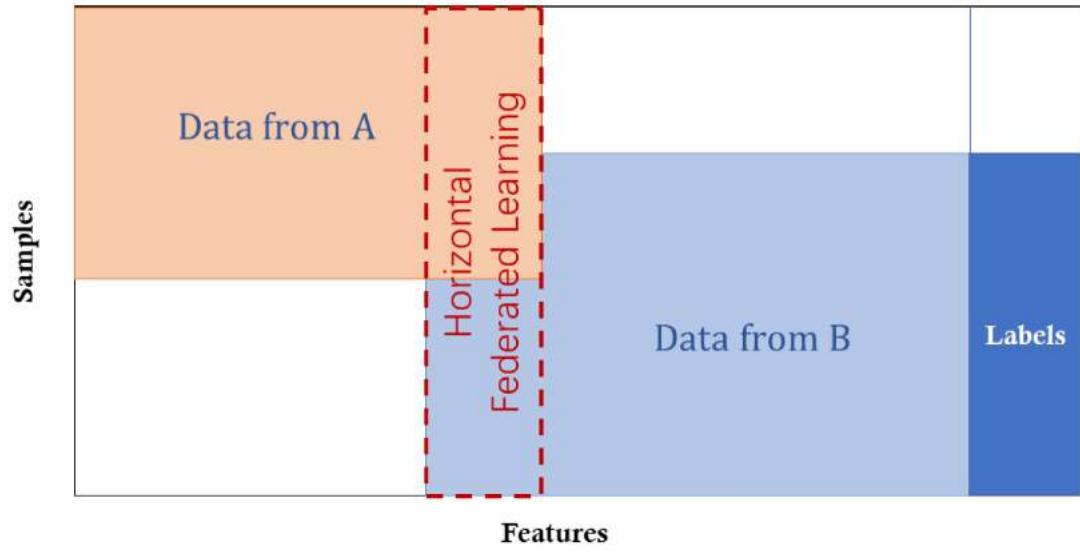
Large overlap of **features** of the two data sets

**Vertical FML**



Large overlap of **sample IDs (users)** of the two data sets

# Horizontal Federated Learning: Divide by Users



(a) Horizontal Federated Learning

**Step 1:** Participants compute training gradients locally

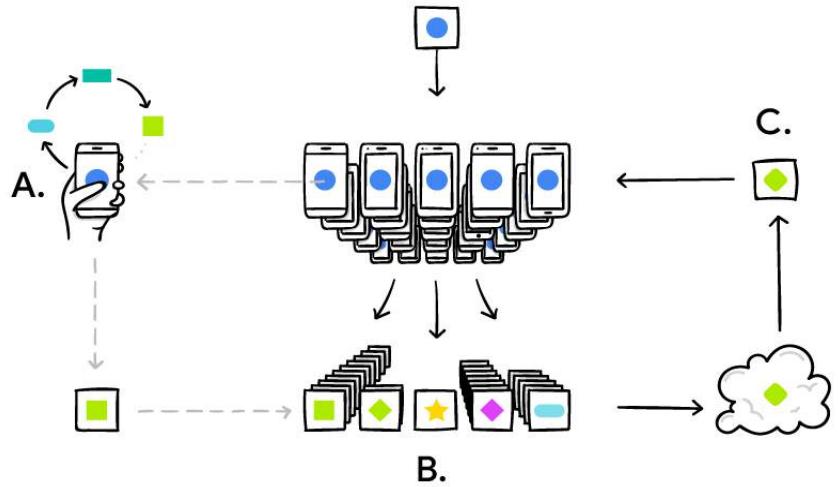
- mask gradients with encryption, differential privacy, or secret sharing techniques
- all participants send their masked results to server

**Step 2:** The server performs secure aggregation without learning information about any participant

**Step 3:** The server sends back the aggregated results to participants

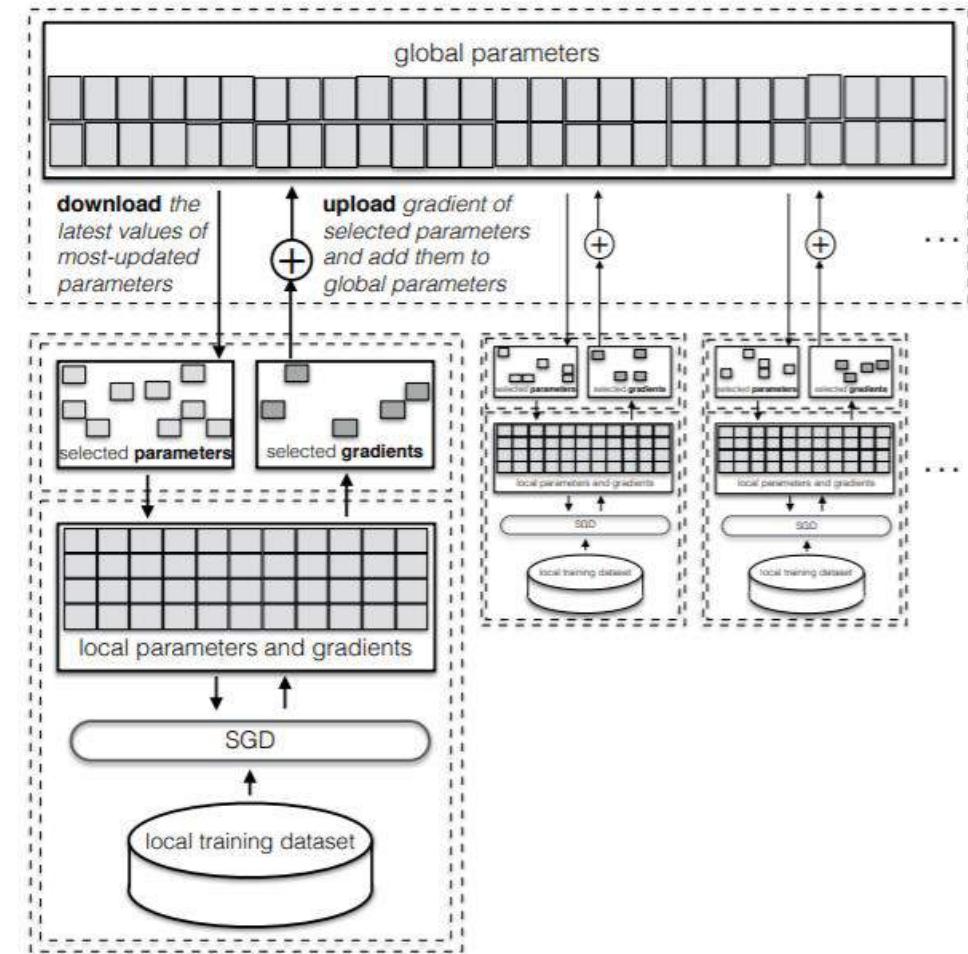
**Step 4:** Participants update their respective model with the decrypted gradients

# Horizontal Federated Learning



H. Brendan McMahan et al, *Communication-Efficient Learning of Deep Networks from Decentralized Data*, Google, 2017

- Multiple clients, one server
- Data is horizontally split across devices, homogeneous features
- Local training
- Selective clients



Reza Shokri and Vitaly Shmatikov. 2015. *Privacy-Preserving Deep Learning*. In Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15). ACM, New York

# Vertical Federated Learning

## Objective:

- Party (A) and Party (B) co-build a FML model

## Assumptions :

- Only one party has label Y
- Neither party wants to expose their X or Y

## Challenges:

- Parties with only X cannot build models
- Parties cannot exchange raw data by law

## Expectations :

- Data privacy for both parties
- model is LOSSLESS

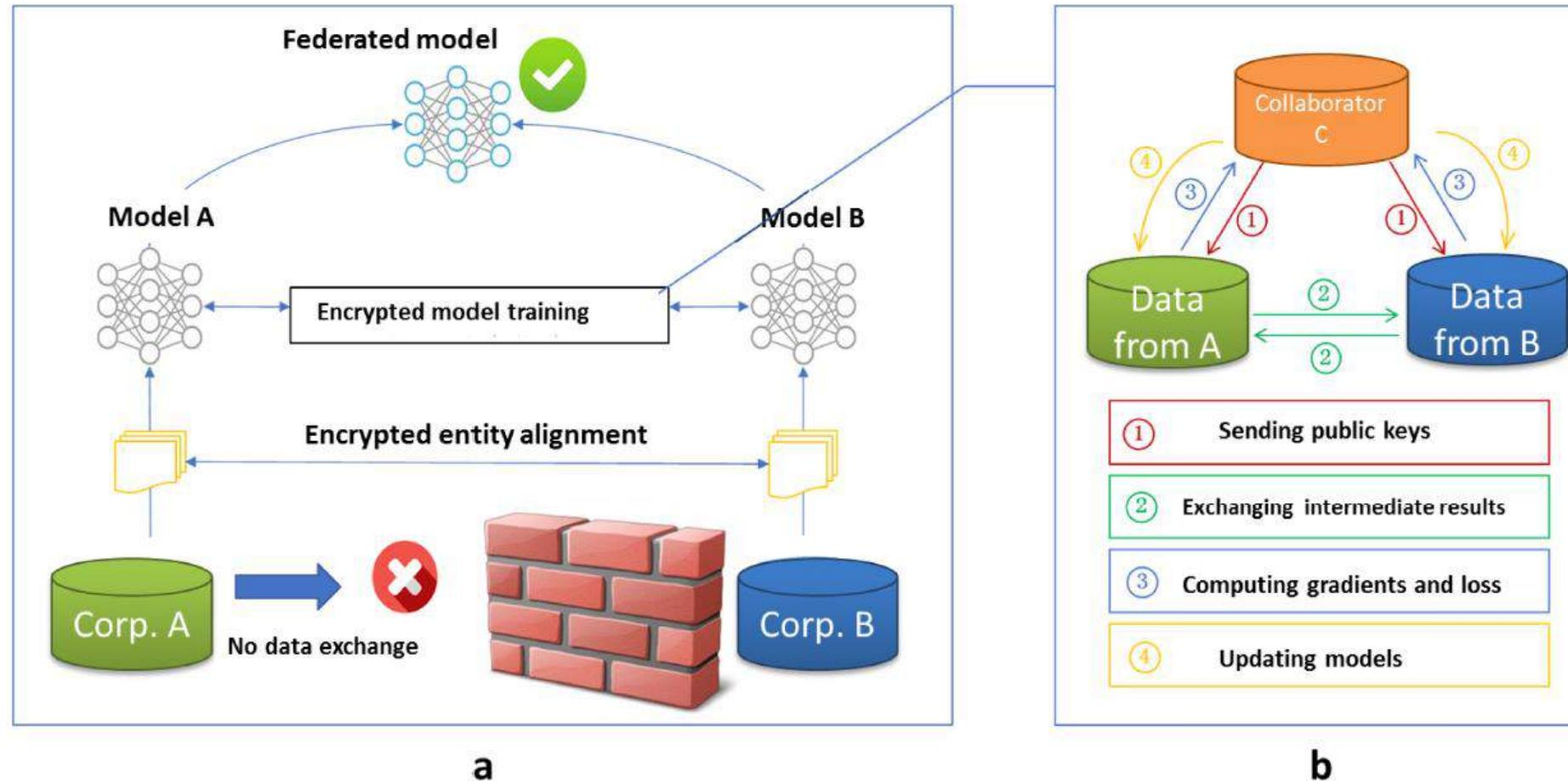


ID	X1	X2	X3	ID	X4	X5	Y
U1	9	80	600	U1	6000	600	No
U2	4	50	550	U2	5500	500	Yes
U3	2	35	520	U3	7200	500	Yes
U4	10	100	600	U4	6000	600	No
U5	5	75	600	U8	6000	600	No
U6	5	75	520	U9	4520	500	Yes
U7	8	80	600	U10	6000	600	No

**Retail A Data**

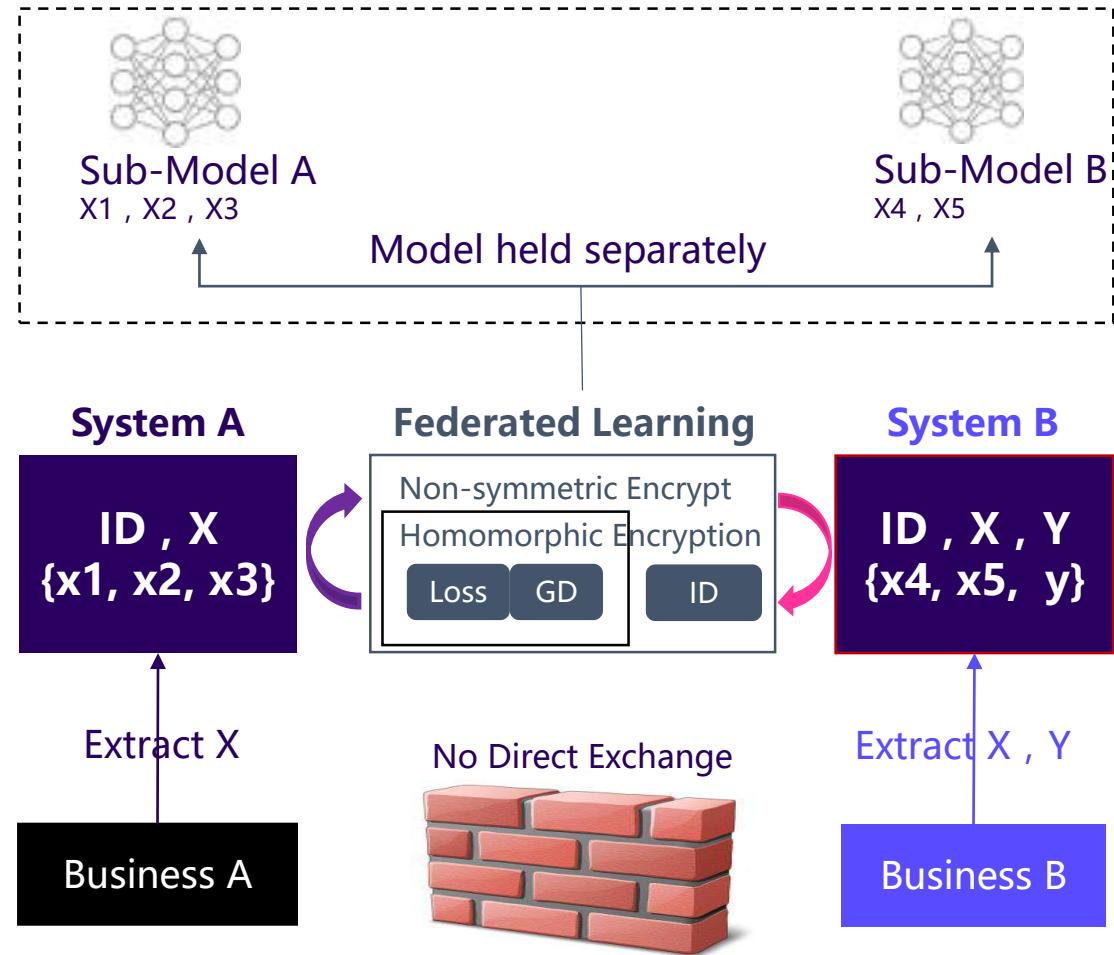
**Bank B Data**

# Vertical Federated Learning

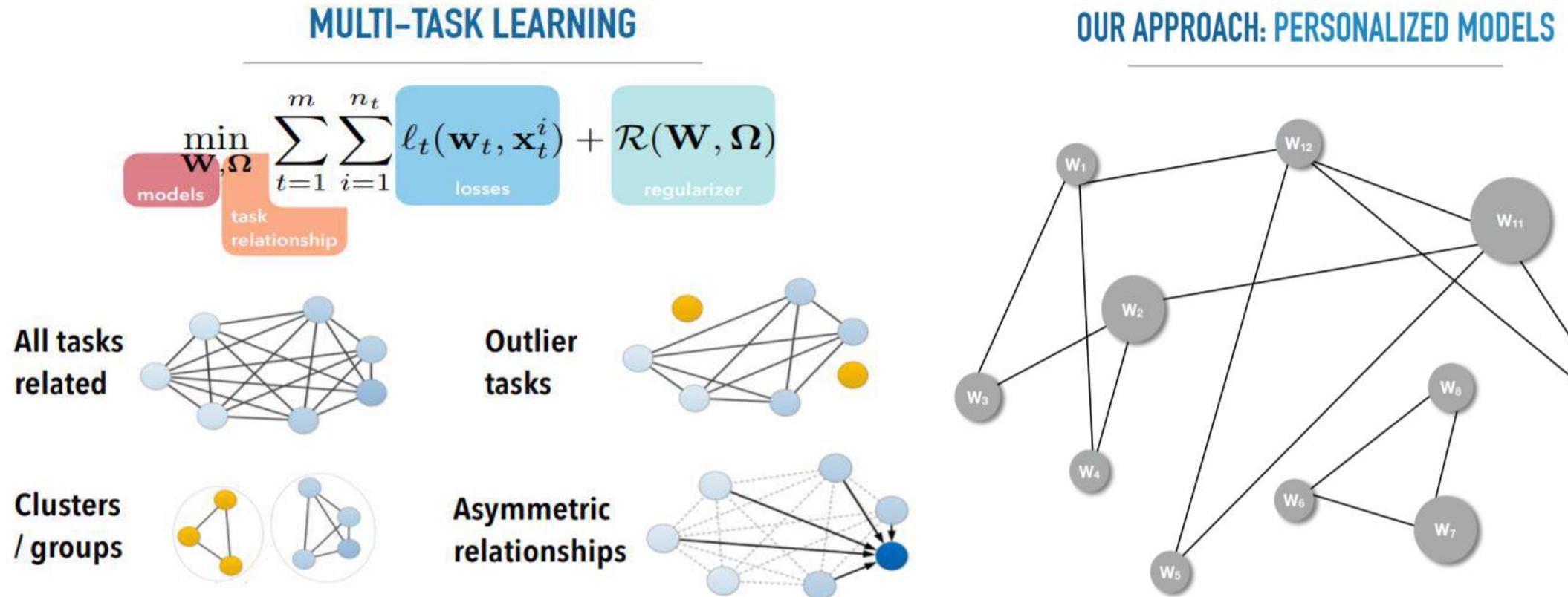


# Vertical Federated Transfer Learning : Features

- **Data Protection :**
  - No different sample ID set leaked
  - No (X, Y) leaked
- **Parameter Protection :**
  - Separately held , jointly used
- **Result :**
  - A has Model<sub>A</sub>
  - B has Model<sub>B</sub>
  - Both models are better than learned separately
- **Property: Lossless**



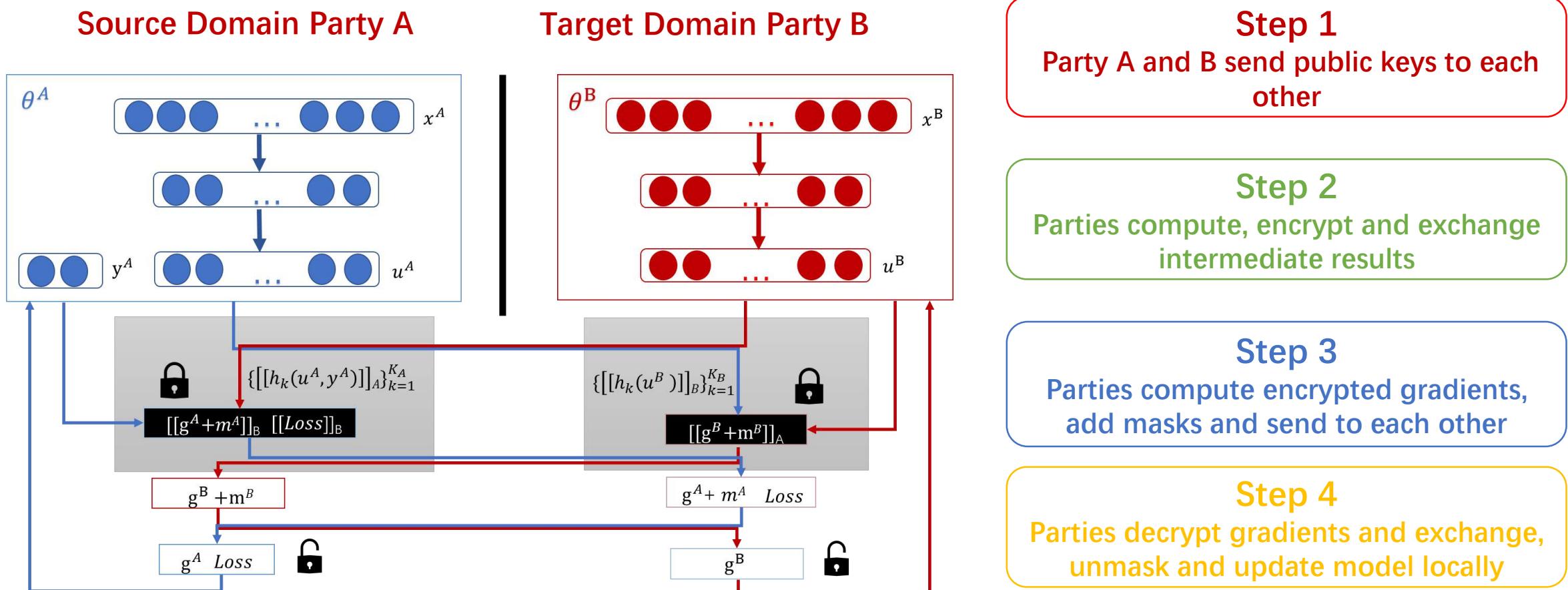
# Federated multi-task learning



Virginia Smith et al. 2017. Federated Multi-Task Learning. In Advances in Neural Information Processing Systems

**How to perform transfer learning  
without sharing data ?**

# Federated Transfer Learning

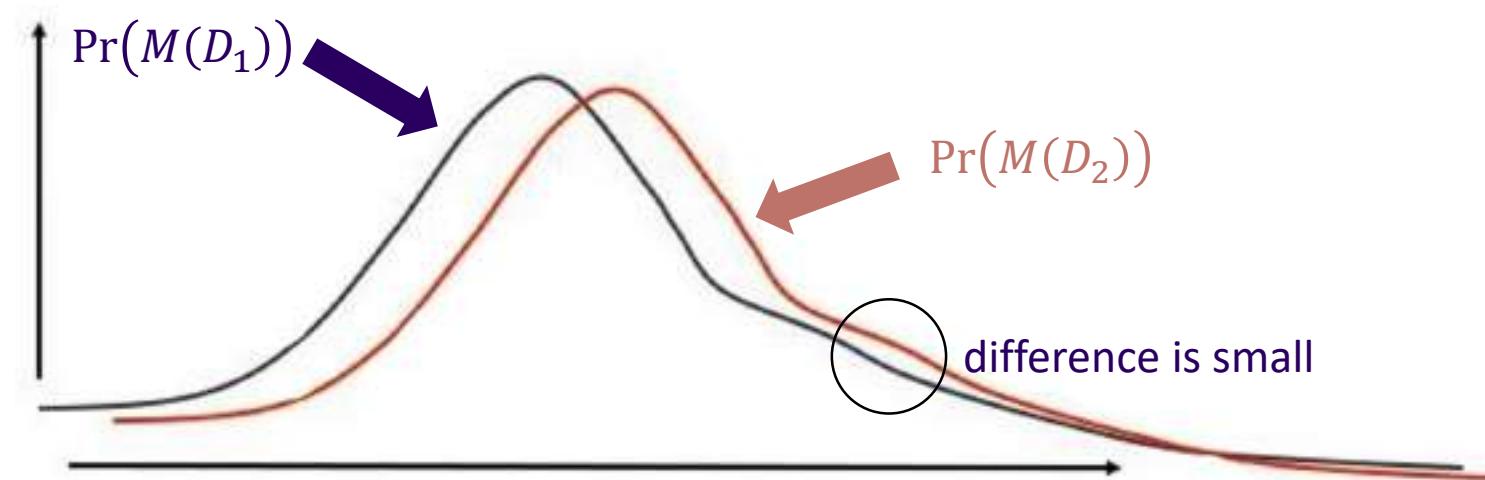


# Differential Privacy: changes in the distribution is too small to be perceived with variations on a single element.

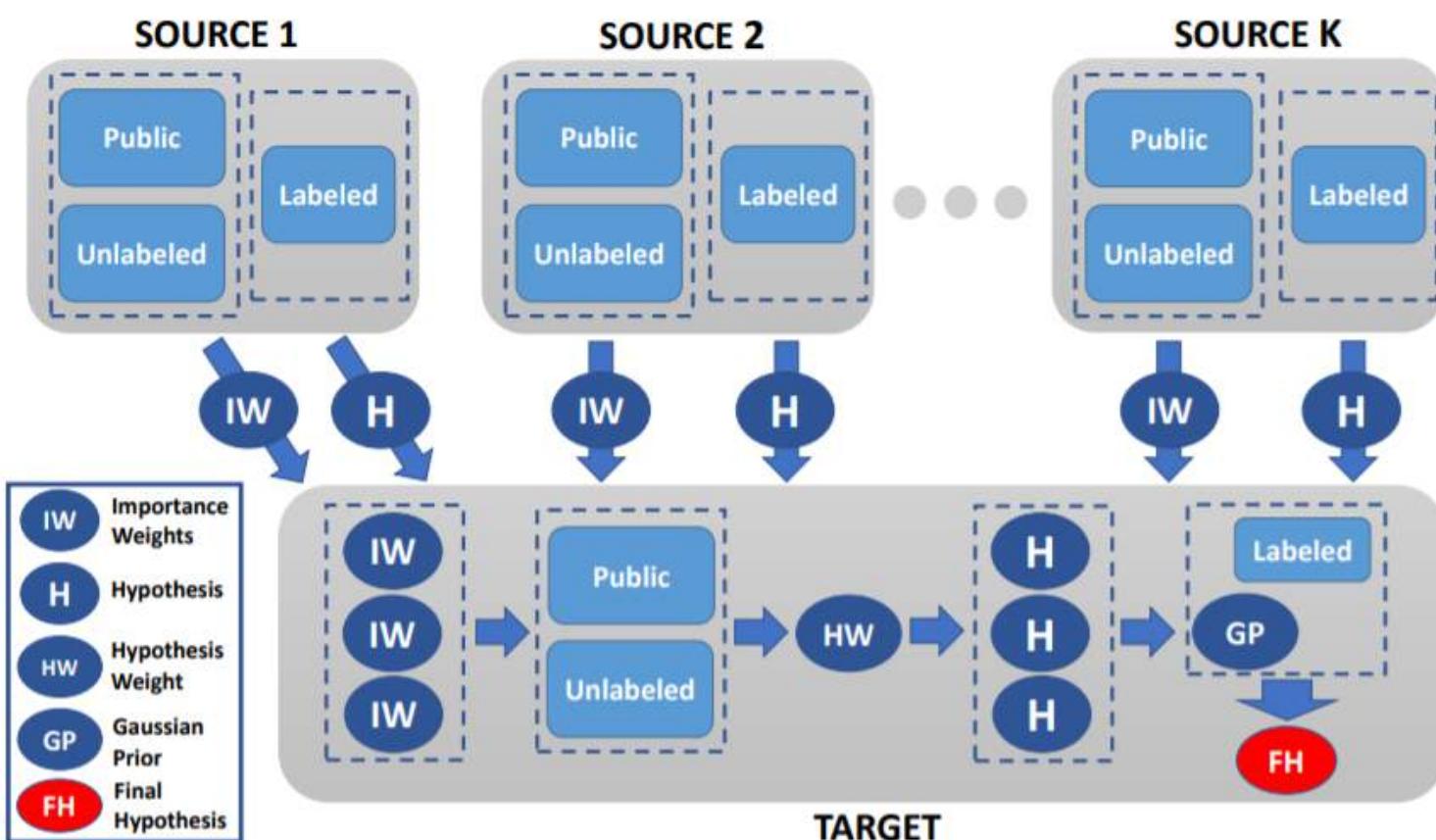
**Definition: Differential Privacy (DP)** [Dwork et.al. 2006, Dwork 2008]

A randomized mechanism  $M$  is  $\epsilon$ -differentially private, if for all output  $t$  of  $M$ , and for all databases  $D_1$  and  $D_2$  which differ by at most one element, we have

$$\Pr(M(D_1) = t) = e^\epsilon \Pr(M(D_2) = t).$$

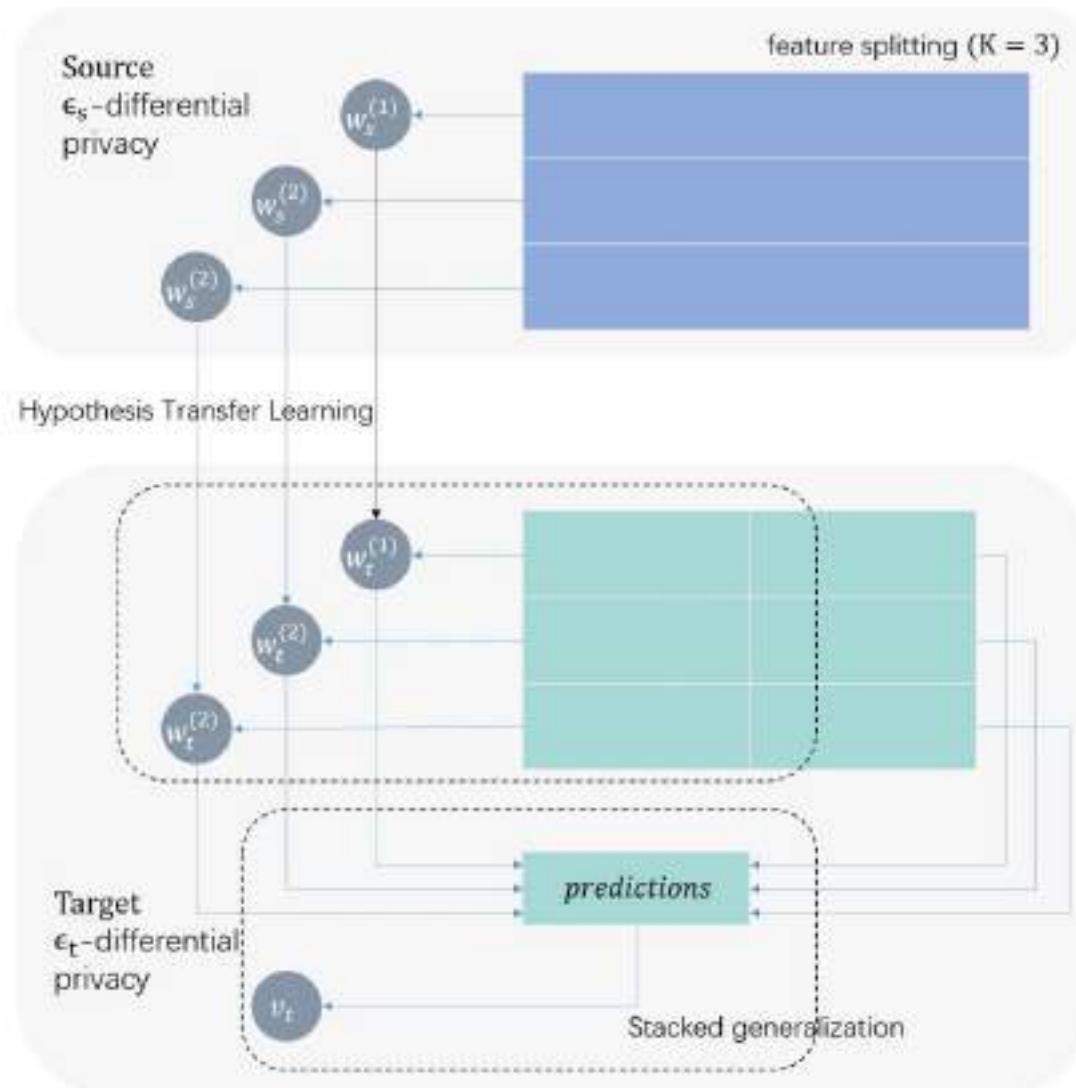


# Differential Privacy with Transfer learning (I)



- leveraging the relatively abundant supply of unlabeled samples and an auxiliary public data set;
- derive the relationship between sources and target in a privacy-preserving manner.
- Source model hypothesis is also differential private

# Privacy-preserving Hypothesis Transfer with feature splitting



- PRL in source

$$\mathbf{w}_s^* = \operatorname{argmin}_{\mathbf{w}} F(\mathbf{w}_s; D, \mathbf{b}, \Delta) + \lambda_s g_s(\mathbf{w}_s)$$

- HTL + PRL in target

$$\mathbf{w}_t^* = \operatorname{argmin}_{\mathbf{w}} F(\mathbf{w}_t; D, \mathbf{b}, \Delta) + \lambda_t g_t(\mathbf{w}_t) + \frac{1}{2} \|\mathbf{w}_t - \mathbf{w}_s^*\|_2^2$$

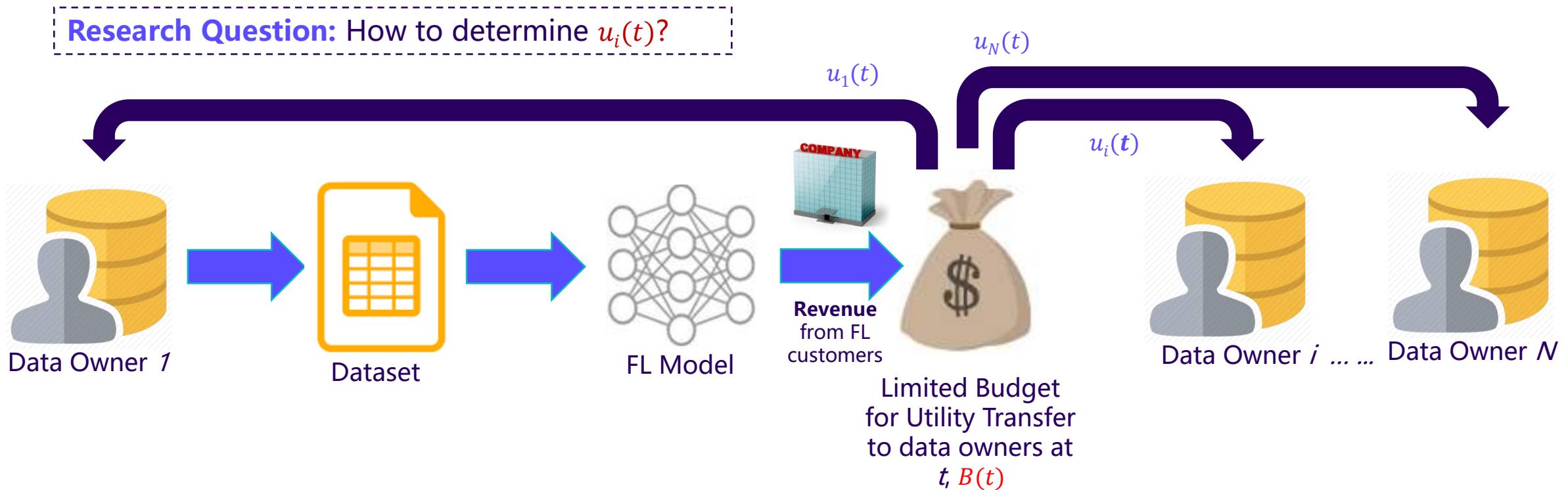
Split features randomly in source and make an ensemble of them in the target, Importance assigned with less noise

$\epsilon$ -differential privacy is guaranteed for both source and target

Privacy-preserving Transfer Learning for Knowledge Sharing. Guo & Yang et.al. Arxiv 1811.09491. 2018

# Incentivize Parties to Join: Federated Learning Exchange

- **Observation:** The success of a federation depends on data owners to share data with the federation
- **Challenge:** How to motivate continued participation by data owners in a federation?



# Federated Learning Exchange

- Objective function for the Federated Learning Exchange (FLE) payoff sharing scheme:

*Maximize* collective utility while *minimizing* inequality among data-owner regrets & waiting times

Maximize:  $\omega U - \Delta$  Regularization weight term

s. t.:

$$\sum_{i=1}^N \hat{u}_i(t) \leq B(t), \forall i, t$$

The actual payoff instalment for  $i$  at  $t$  if the federation does not have enough budget to pay out the full incentive amount for all data owners in one go

Solution:

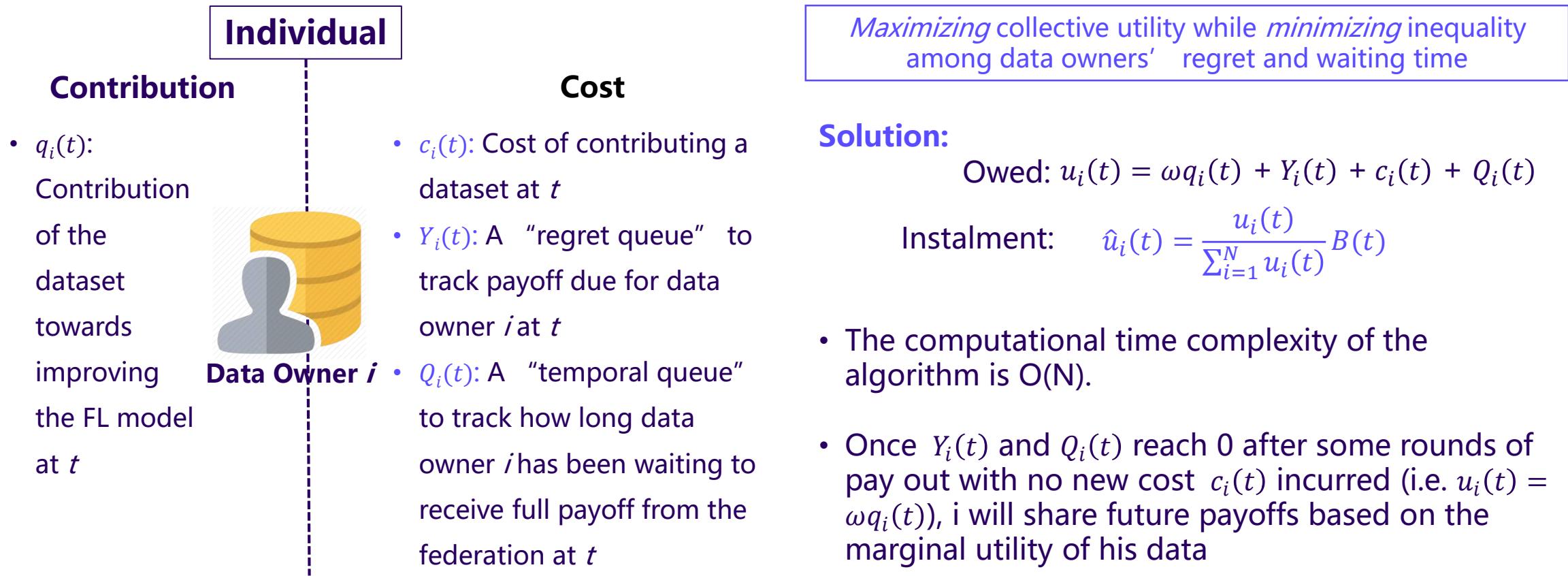
Owed:  $u_i(t) = \omega q_i(t) + Y_i(t) + c_i(t) + Q_i(t)$

Instalment:  $\hat{u}_i(t) = \frac{u_i(t)}{\sum_{i=1}^N u_i(t)} B(t)$

- The computational time complexity of the algorithm is  $O(N)$ .
- Once  $Y_i(t)$  and  $Q_i(t)$  reach 0 after some rounds of pay out with no new cost  $c_i(t)$  incurred (i.e.  $u_i(t) = \omega q_i(t)$ ),  $i$  will share future payoffs based on the quality of his data contribution.

# FL Payoff-Sharing

- In order to fully commercialize federated learning among different organizations, a fair platform and incentive mechanisms need to be developed



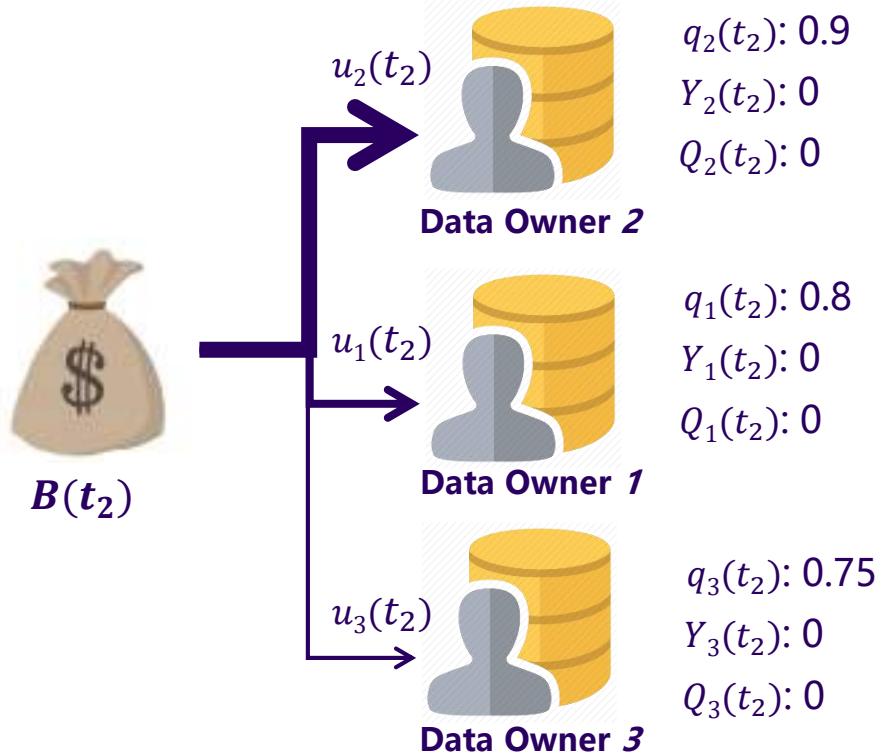
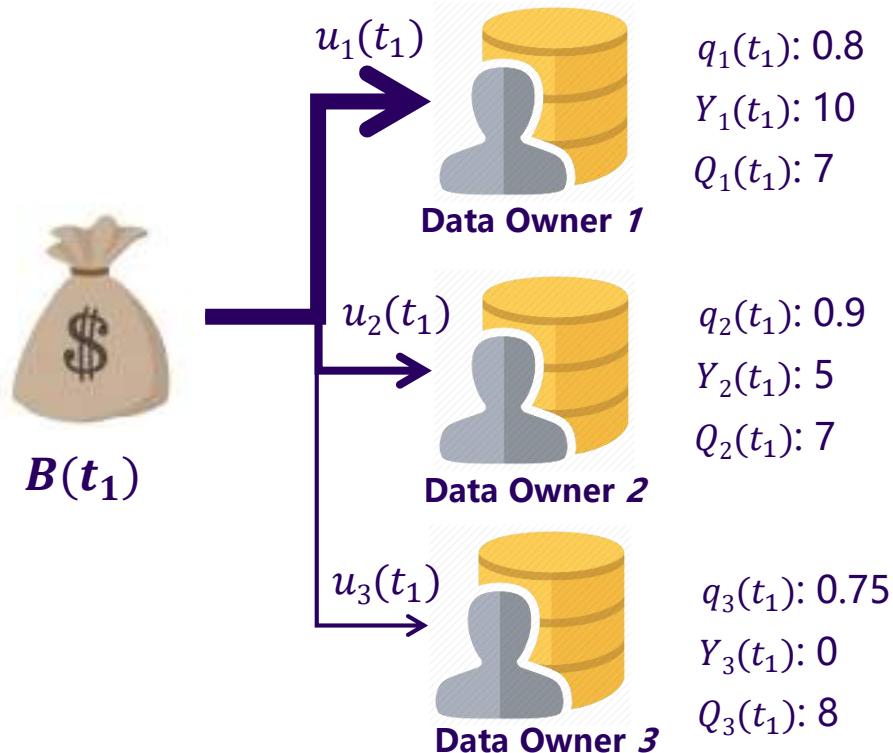
# Federated Machine Learning: Advantages

	Coalition games with transferable utility [1]	Labour union games [2,3]	Fair-value games / Shapley games [2,3]	Federated Learning (this work)
Players do not need to engage in complex negotiations	X	√	√	√
Payers can join multiple coalition at the same time, cost for joining a coalition, the value of a dataset does not depreciate after being shared, and players' time spent waiting for cost to be compensated	X	X	X	√
Players' marginal contribution is important	√	X	√	√

- 1) B. Faltings, G. Radanovic & R. Brachman. *Game theory for data science: Eliciting truthful information*. Morgan & Claypool Publishers, p. 152, 2017.
- 2) J. Augustine, N. Chen, E. Elkind, A. Fanelli, N. Gravin & D. Shiryaev. Dynamics of profit-sharing games. *Internet Mathematics*, 1:1–22, 2015.
- 3) S. Gollapudi, K. Kollias, D. Panigrahi & V. Pliatsika. Profit sharing and efficiency in utility games. In *ESA*, pp. 1–16, 2017.

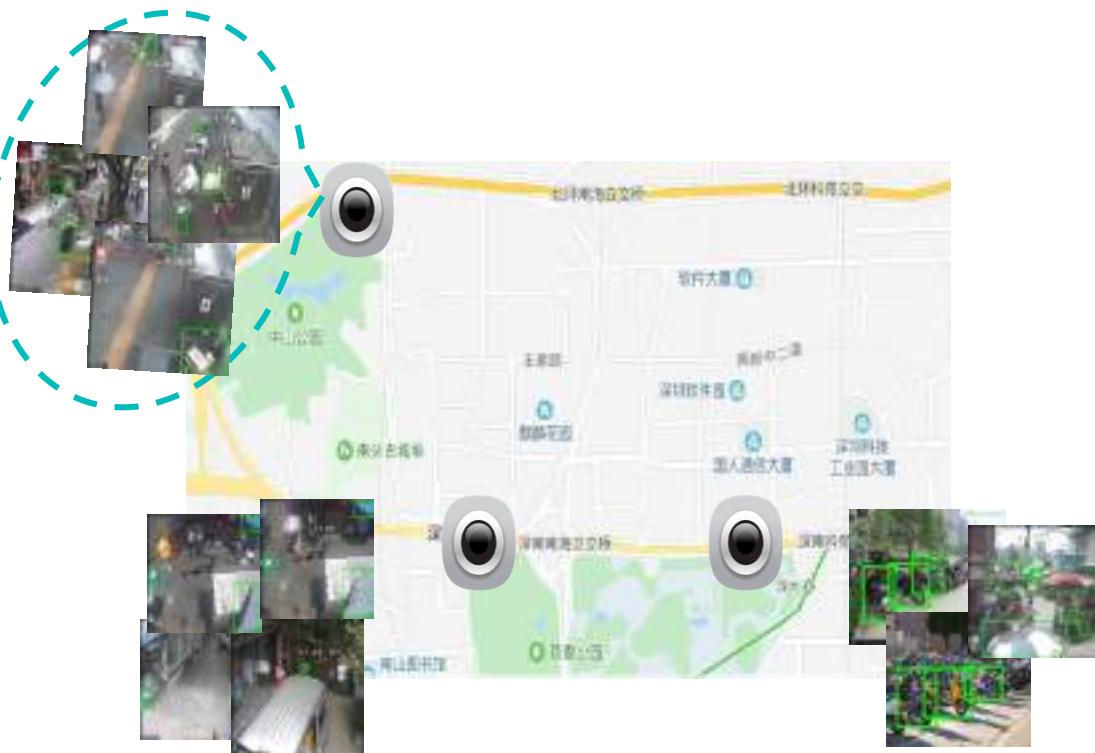
# Examples

We set  $\omega = 10$  in the following examples



# Application: FML Network for Object Detection

- For parking and street vendor violation
- A partner project of Webank AI and Extreme Vision in Shenzhen, China



## Challenges

- Difficult detection task with few labels;
- Data are scattered; Expensive to centralize and manage data;
- Delayed feedback and delayed model updates.

## A federated learning approach

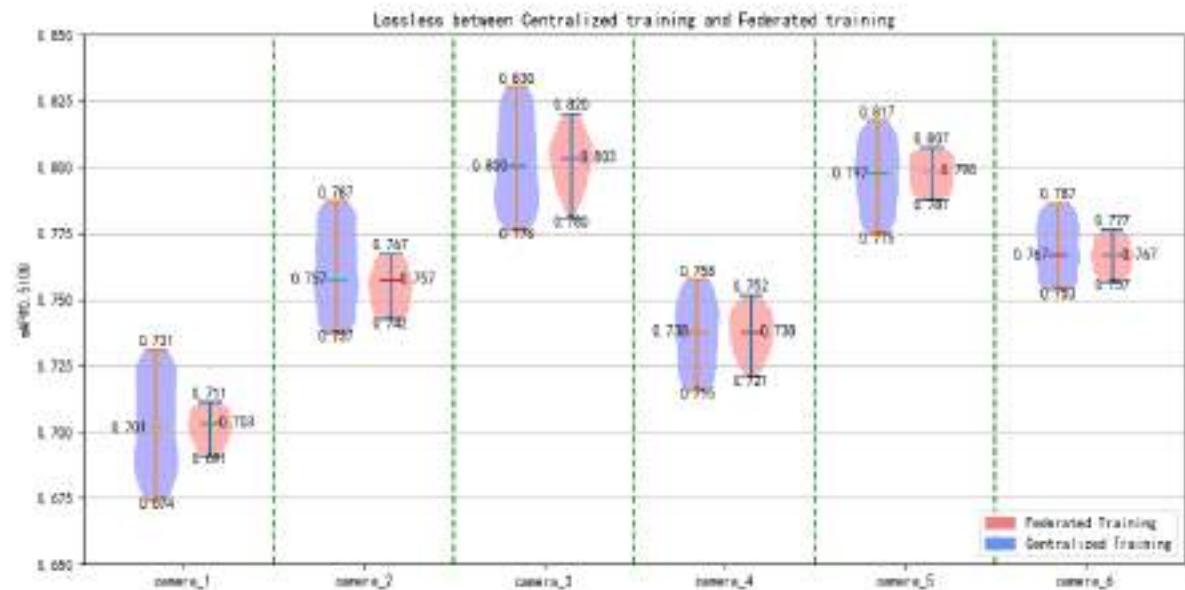
- Online feedback loop and model updates;
- No need to upload and centralize data;
- No sharing data.

7 class : {table, chair, carton, sunshade, basket, gastank, electromobile} with 6 cameras, 1922 images

**Federated model improved local model by 15%**



**lossless performance  
(Centralized model vs federated model)**



# Federated AI Ecosystem



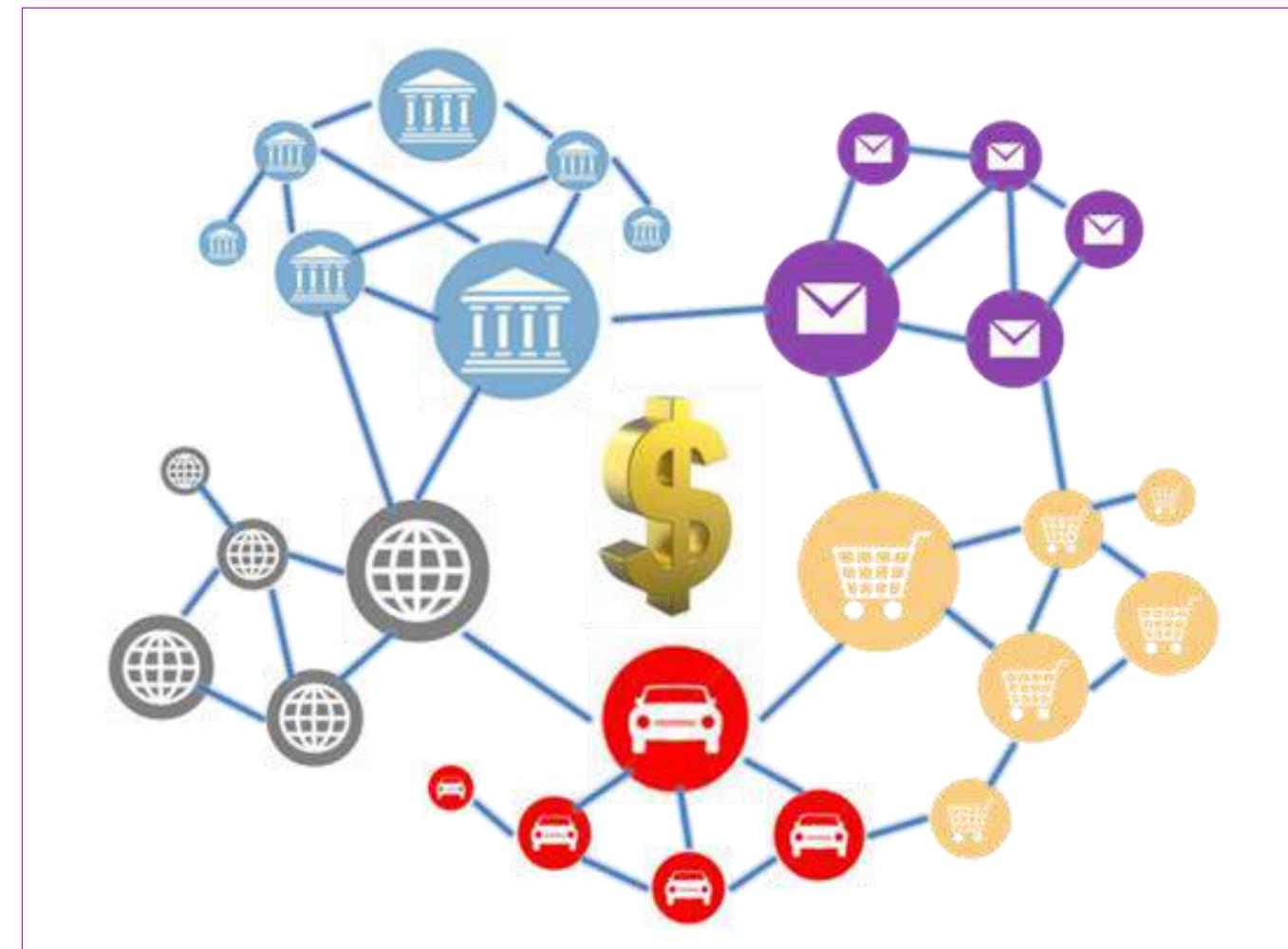
Security & Compliance



Lossless performance



Effective for small data and weak-supervision problem



# IEEE Standard P3652.1 – Federated Machine Learning

## ➤ Title:

- Guide for Architectural Framework and Application of Federated Machine Learning
- Description and definition of federated learningThe types of federated learning and the application scenarios to which each type applied
- Performance evaluation of federated learning
- Associated regulatory requirements

## ➤ First working group meeting:

- First working group meeting
- Dates: February 21~22, 2019
- Location: Shenzhen, China
- <https://sagroups.ieee.org/3652-1/>

# Open Source in Feb 2019 – Federated AI Technology Enabler (FATE)

- **FATE is an open-source project initiated by Webank's AI Department**
  - Supports federated learning architectures including horizontal federated learning, vertical federated learning and federated transfer learning
  - Implements secure computation protocols based on homomorphic encryption and multi-party computing ( MPC )
  - Supports the secure computation of various machine learning algorithms, including logistic regression, tree-based algorithms, and deep learning and transfer learning



# Federated AI Ecosystem

Collaborative Learning and Knowledge Transfer Preserving Data Privacy and Confidentiality

[GitHub](#)[White Paper](#)

**Find more information at <https://www.fedai.org/>**

# Summary

- AI' s Data Challenge: data shortage, regulations, and fragmentation
- Transfer Learning: from pretrained large models to small data
- Federated Machine Learning: secure collaboration in model building
- Federated Transfer Learning, Incentive Mechanisms and Open Source Frameworks
- <https://sagroups.ieee.org/3652-1/>
- <https://www.fedai.org/>