

Effective Knowledge Representation and Utilization for Sustainable Collaborative Learning across Heterogeneous Systems

Trong Nghia Hoang

School of Electrical Engineering and Computer Science
Washington State University
trongnghia.hoang@wsu.edu

Abstract

The increasingly decentralized and private nature of data in our digital society has motivated the development of collaborative intelligent systems that enable knowledge aggregation among data owners. However, collaborative learning has only been investigated in simple settings. For example, clients are often assumed to train solution models *de novo*, disregarding all prior expertise. The learned model is typically represented in task-specific forms that are not generalizable to unseen, emerging scenarios. Finally, a universal model representation is enforced among collaborators, ignoring their local compute constraints or input representations. These limitations hampers the practicality of prior collaborative systems in learning scenarios with limited task data that demand constant knowledge adaptation and transfer across information silos, tasks, and learning models, as well as the utilization of prior solution expertise. Furthermore, prior collaborative learning frameworks are not sustainable on a macro scale where participants desire fairness allocation of benefits (e.g., access to the combined model) based on their costs of participation (e.g., overhead of model sharing and training synchronization, risk of information breaches etc.). This necessitates a new perspective of collaborative learning where the server not only aggregates but also conducts valuation of the participant's contribution, and distribute aggregated information to individuals in commensurate to their contribution. To substantiate the above vision, we propose a new research agenda on developing effective and sustainable collaborative learning frameworks across heterogeneous systems, featuring three novel computational capabilities on knowledge organization: model expression, comprehension and valuation.

1 Introduction

Modern problem-solving systems are frequently integrated into a complex and diverse information network. For instance, think of a smart health monitoring system, e.g. (Hasantabar et al. 2020), that compiles and disseminates analytical findings derived from a wide range of patient data. This data is often scattered across various hospitals, clinics and numerous personal medical wearable devices. All of which may have unique ownership, hardware setups (e.g., compute and communication bandwidths), data distributions (e.g., patient demographics) and collection mecha-

nisms (e.g., which signals are being monitored). We refer to these different aspects as system heterogeneities.

As these private data sources are not centrally owned, privacy-preserving and privacy-compliant knowledge sharing technologies have become basic requirements to facilitate AI collaboration. To this end, federated learning (FL) proposes to combine information from private data silos after they have been distilled into a common representation, or more simply the weights of a learning model (Konečný et al. 2016; McMahan et al. 2017). However, there are many gaps between the vanilla FL setting and its practical use cases. For instance, the shared learning model might not simultaneously fit the various local client constraints (e.g., compute capacities and input representations). Further, FL is assumed to train solution models from scratch and does not build on prior knowledge. Learned knowledge might also be represented in task-specific forms (e.g., models trained on specific data), and is consequently not generalizable or adaptable to unseen task scenarios.

The limitations above do not align with the requirements of continual collaborative learning, which demands adaptability and transferability of knowledge across various information silos, tasks, and learning models. For example, consider the smart health monitoring scenario above where personal devices summarize patient statistics for diagnosis and treatment. Patients use various monitoring devices, requiring the system to handle diverse data representations. The prediction models must also be adaptable to new, emerging diseases with limited data. This is not possible without a representation that can robustly accommodate such system and data heterogeneities, as well as a corresponding resource-aware knowledge aggregation mechanism that can consolidate, leverage and adapt prior solution expertises to combat data scarcity in new task context. Furthermore, most existing works in collaborative learning have not considered its sustainability on a macro scale of an ecosystem, featuring a large, ever-growing crowd of continual or lifelong learning agents in business competitive scenarios such as smart health monitoring where information sharing essentially requires cooperation across different monitoring platforms developed by different for-profit organizations. For such learners, their key concern is whether the benefit (e.g., access to a better combined model) derived from the collaboration out-

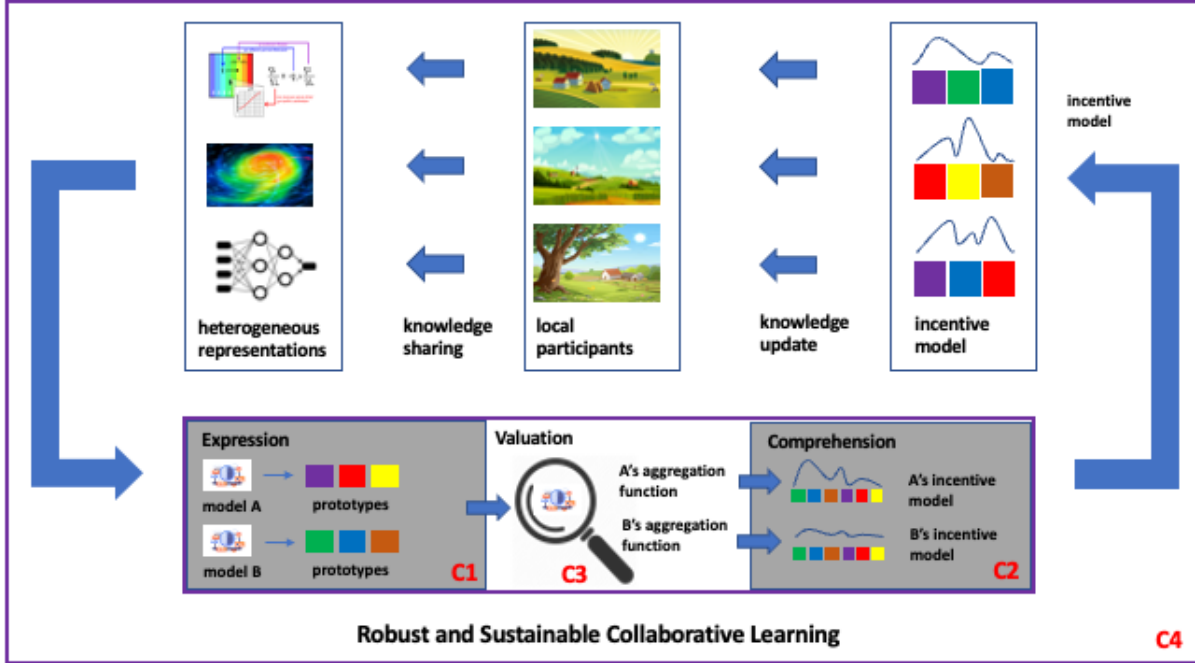


Figure 1: Overview of our proposed collaborative learning framework with key research thrusts **C1**, **C2**, **C3** and **C4** highlighted.

weighs the cost of participation (e.g., overhead of training synchronization and/or model sharing, risk of private information breaches etc.). To mitigate this concern, an appropriate incentivization mechanism is needed to ensure fair benefit distribution and hence, sustained engagement from the participants.

To bridge this gap between federated learning and its potential application to such continual collaboration scenarios, we envision a task-agnostic and resource-aware framework for model representation and aggregation. Within this framework, pre-trained models capturing related expertises are broken down into a diverse set of task-agnostic functions, each associated with distinct task embedding patterns. Local solution models can then be represented using subsets of these functions, selected based on their proximity to the task data. This approach transforms model aggregation into a set summarization problem, affording it the desirable adaptability, transferability, and compositionality across different information silos and tasks with different model and data representations. The current progress and specific directions of the envisioned research are detailed next.

2 Robust Collaborative Learning

Our proposed research agenda will focus on developing several key computational capabilities, namely **expression**, **comprehension** and **valuation**, that enable effective communication among collaborative learners with heterogeneous models or knowledge representations. First, **expression** pertains to a learning collaborator’s capacity to de-

termine what information to communicate to facilitate the global decomposition of local models into reusable patterns. On the other hand, **comprehension** focuses on the ability to grasp the semantic associations and compositionality of these patterns. Last but not least, **valuation** grants participants clarity to assess whether the benefit derived from the collaboration (e.g., model utility) is worth the cost of participation (e.g., overhead of model sharing and training synchronization, risk of information breaches, etc.). This is essential to ensure sustained engagement of learning participants through proper incentivization. For example, based on such fair valuation mechanism, learning agents that provide more useful information should be allocated more bandwidth to access the combined model than others. Such capabilities on model **expression**, **comprehension**, and **valuation** will form the backbone to a sustained collaborative learning ecosystem in large scale. This vision will be substantiated by the following research agenda. Its overall workflow is depicted in Fig. 1.

C1. Model Expression: We will focus on devising effective algorithms that factorize pre-trained black-box models into a set of task-agnostic and comprehensible predictors called *prototypes* (Hoang et al. 2019a,b, 2020; Lam et al. 2021). This allows us to represent prior problem-solving expertise in a modular and transferable fashion, where distinct knowledge patterns are captured by these context-independent model prototypes and can be recombined to synthesize novel solutions to new task contexts. This will re-imagine the existing paradigm of model fine-tuning, producing models that

are both effective and interpretable when adapting to new tasks, which is crucial in pre-training and fine-tuning with foundation or large pre-trained models.

One potential direction, as previously investigated by Hoang et al. (2020), is to find a task-agnostic embedding of the pre-trained models $\mathbf{B}_{\tau_1}, \mathbf{B}_{\tau_2}, \dots, \mathbf{B}_{\tau_p}$ on a factorized (latent) space $\mathbf{H} = \mathbf{W} \times \mathbf{Z}$ where \mathbf{Z} encodes task-agnostic concepts that underline the black-box’s inferential mechanism while \mathbf{W} isolates generic input patterns from \mathbf{Z} . This is achieved via using a probing, unlabeled dataset \mathbf{U} to sample the inferential patterns of the pre-trained models, which are expressed in terms of a collection of triplets (\mathbf{x}, η, τ) . Here, \mathbf{x} and η denote the corresponding input and (soft) output of \mathbf{B}_τ while τ encodes the task information (e.g., one-hot vectors representing task identities or context prompt engineered by domain experts).

As a concrete example, consider the problem of handwritten digits classification, we want to encourage the following behavior in our model: \mathbf{z} encodes the information central to making predictions (i.e. the numerical value of the digit) whereas \mathbf{w} encodes information that does not influence the prediction, such as the width and tilt of the strokes, the light intensity of the images, and other abstract stylistic properties. We want to find an embedding of such probing data on the aforementioned factorized space $\mathbf{H} = \mathbf{W} \times \mathbf{Z}$ from which task-agnostic prototypes can be derived.

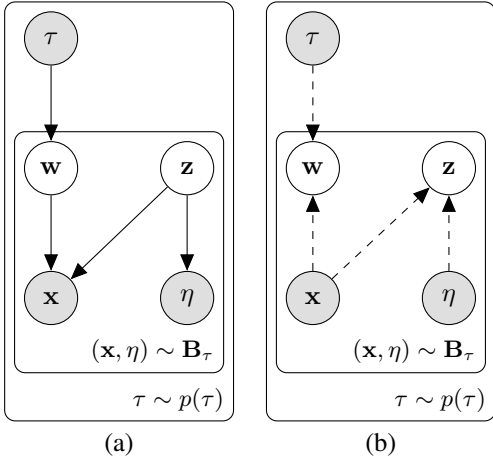


Figure 2: Graphical models of (a) the generative and (b) inference networks – $p(\mathbf{w}, \mathbf{z}, \mathbf{x}, \eta, \tau; \theta, \gamma, \alpha)$ and $q(\mathbf{w}, \mathbf{z}|\mathbf{x}, \eta, \tau; \phi)$, respectively – in our model embedding framework. The dashed arrows in (b) indicate the posterior surrogates that form the inference network.

Under this modeling paradigm (see Fig. 2a), we adopt the following parameterization for $p(\mathbf{w}, \mathbf{z}, \mathbf{x}, \eta, \tau; \theta, \gamma, \alpha)$,

$$p(\mathbf{w}, \mathbf{z}, \mathbf{x}, \eta, \tau; \theta, \gamma, \alpha) \triangleq \begin{aligned} & p_\theta(\mathbf{x}|\mathbf{w}, \mathbf{z})p_\gamma(\mathbf{w}|\tau) \\ & p_\alpha(\eta|\mathbf{z})p(\tau)p(\mathbf{z}), \end{aligned} \quad (1)$$

where θ, γ, α denote an abstract parameterization often implemented in form of a (deep) neural network. Thus, learning this representation means learning (θ, γ, α) that best ex-

plains the observations (\mathbf{x}, η, τ) which were collected by observing the prediction of $\mathbf{B}_{\tau_1}, \mathbf{B}_{\tau_2}, \dots, \mathbf{B}_{\tau_p}$ at the unlabeled data $\mathbf{x} \in \mathbf{U}$. This is often learned via a variant of variational auto-encoder (VAE) (Kingma and Welling 2013) which exploits a parameterized inference network (see Fig. 2b),

$$q_\phi(\mathbf{w}, \mathbf{z}|\mathbf{x}, \tau, \eta) \triangleq q_\phi(\mathbf{z}|\mathbf{x}, \eta) q_\phi(\mathbf{w}|\mathbf{x}, \tau), \quad (2)$$

to define and maximize a variational lower-bound of the probing data’s likelihood. The learned generative model can then be used to express any pre-trained model $\mathbf{B}_\tau \equiv p(\eta | \mathbf{x}, \tau)$ in terms of an integration over a spectrum of task-agnostic prototypes via the fundamental laws of probability,

$$p(\eta|\mathbf{x}, \tau) \propto \mathbb{E}_{\mathbf{w} \sim p_\gamma(\mathbf{w}|\tau)} \left[g_{\mathbf{w}}(\eta|\mathbf{x}; \theta, \alpha) \right], \quad (3)$$

where $g_{\mathbf{w}}(\eta|\mathbf{x}; \theta, \alpha)$ denote a task-agnostic prototype, which is expressed as an integration over \mathbf{z} ,

$$g_{\mathbf{w}}(\eta|\mathbf{x}; \theta, \alpha) = \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} \left[p_\theta(\mathbf{x}|\mathbf{w}, \mathbf{z})p_\alpha(\eta|\mathbf{z}) \right], \quad (4)$$

which can be synthesized to solve an unseen task τ_* ,

$$\mathbf{B}_{\tau_*}(\mathbf{x}) = \mathbb{E}_{\mathbf{w} \sim p_\gamma(\mathbf{w}|\tau_*)} \left[\arg \max_\eta g_{\mathbf{w}}(\eta|\mathbf{x}; \theta, \alpha) \right]. \quad (5)$$

Nonetheless, while Eq. (5) can be approximated reasonably well following the preliminary experiments in (Hoang et al. 2020), its complexity will explode exponentially in the size of the latent coordinates \mathbf{w} , which is however essential to embed pre-trained models with highly sophisticated parameterization on complex data spaces.

A potential approach to mitigate this is to generalize the aforementioned technique to the context of prompt-based or adapter-based fine-tuning framework which interestingly represents customized solutions in terms of a set of (learnable) input prefixes (Wang et al. 2021) to a transformer-based pre-trained model or a set of light-weight, low-complexity neural adapters (Hu et al. 2021) used to replace the pre-trained computation workflow of some existing neural blocks (e.g., the pre-trained model’s prediction head). Alternatively, we will also investigate more direct approaches that leverage pre-trained models’ outputs on target inputs and/or their (learnable) ensemble to enable in-context learning or soft-prompt generator for existing prompt-tuning framework, following our recent findings in few-shot learning with black-box ensemble (Hoang and Hoang 2024).

C2. Model Comprehension: To complement the above research thrust, model comprehension focuses on developing statistical techniques to associate, align and re-combine relevant prototypes extracted from different pre-trained models (Yurochkin et al. 2019a; Hoang et al. 2020; Lam et al. 2021) to solve new tasks. This is a key issue in continual and/or federated learning scenarios with data, model and/or system heterogeneities which result in a diverse range of local representations of knowledge. Although the research in **C1** can be repurposed to embed and decompose such heterogeneous knowledge representations into context-independent prototypes, the key issue here is, however, the partial accessibility of local knowledge per learning epoch due to the nature of continual and federated learning.

For example, in continual learning, the learning agent only has access to local data/solution of a single task per step. In cross-device federated learning, only a small subset of learning agents (e.g., local devices) will participate in knowledge aggregation. This will inevitably create asymmetries in the knowledge representation across agents and learning epochs. To mitigate this, the learning agents need to infer the correspondence between knowledge modules that were distilled in different contexts and orders. Addressing this helps re-imagine the knowledge aggregation mechanism in federated and/or continual learning as a distributed or streaming (random) set modeling task. This has been previously investigated in the context of neural networks that decompose into sets of neurons (Yurochkin et al. 2019a,b). Local neurons can then be communicated among learning agents and partitioned into clusters whose centers are leveraged to derive aggregated neurons which can be assembled into a better model. Such approaches are applicable to traditional feed-forward, convolutional and recurrent neural nets but were not designed to work with more recent modern architectures such as self-attention (Vaswani et al. 2017).

The main focus in this thrust is therefore to develop more robust and versatile random set modeling frameworks that are generalizable to modern architectures. For example, multi-head attention in transformer (Vaswani et al. 2017) can be re-characterized in terms of a sparse Gaussian process model (Bui et al. 2024), which are represented by its set of inducing inputs. Our previous work in (Yurochkin et al. 2019b) has in fact shown that it is possible to extend the above neuron clustering scheme to inducing input clustering to aggregate sparse Gaussian processes across local learning agents. This allows for a broader generalization from neuron clustering to prototype clustering, whose characterization is expressed in terms of solutions to certain optimization tasks, which is even more sophisticated. We will further investigate whether such characterization can also be cast into the aforementioned set modeling framework. We envision that the rich literature on probabilistic set modeling frameworks and/or their applications to (streaming) clustering scenarios can be leveraged to drive research in this direction. We will also investigate a more direct generalization of our previous work (Yurochkin et al. 2019a,b) to prompt-tuning context whose solutions are naturally characterized in terms of prompt sets, which are readily integrable into the envisioned set modeling framework.

C3. Model Valuation: The value of knowledge summaries (e.g., how useful they are to others) must be quantifiable, so that individual contribution of participants can be assessed. The aggregated knowledge must be distributed such that benefits are proportionate to individual contributions. Data valuation has previously been proposed to address fairness in federated learning (FL). However, prior literature (Wang et al. 2020; Li et al. 2023; Wei et al. 2020; Fan et al. 2024) has not considered cases in which data appraising techniques need to assess value of data from heterogeneous data summaries (e.g., simulation processes and data-driven models). Most works in this space assume direct access to data or a common model derivative of the data. This assumption is

however impractical in real-world domains that require heterogeneous formulations for different aspects of the same task (Wang et al. 2020; Ghorbani, Kim, and Zou 2020; Ghorbani and Zou 2019; Yoon, Arik, and Pfister 2019; Sim et al. 2023). For instance, in smart farming, spatio-temporal variation models of crop yield prediction might be formulated differently across collaborating farms, such as using biophysical simulation systems or training machine learning models to predict unknown behaviors from collected data. Local participants might also employ different sensor instruments with distinct data types (e.g., photos of crop field, soil pH, atmospheric condition) and resolutions (e.g., density of sensors, measurement precision) which inevitably leads to models with heterogeneous forms (Dhanaraju et al. 2022).

An effective collaborative learning system is therefore centered on an effective knowledge fusion operator and a computationally accessible notion of information value for any aggregable subset of data summaries. This notion must characterize consistent appraisals. That is: (a) a subset of data summaries cannot be more valuable than any of its supersets; and (b) an aggregated model distributed to an agent is at least as valuable as its individual data summaries. Algorithms that compile and generate aggregations of data summaries can be developed based on this notion to guarantee fair incentives for participating individuals.

Our recent investigation (Sim et al. 2023) achieved this using differential privacy (DP) (Mironov 2017) as an incentive. Each participant can select its required DP guarantee and sanitize its sufficient statistic (SS) of the local model with algorithmically crafted noises. The server values such perturbed SS using a notion of Bayesian surprise (Itti and Baldi 2009), which characterizes how much new information a local participant contributes. Intuitively, a higher DP guarantee requires higher order of perturbation which reduces the informativeness of the shared SS. Once the shared SS are aggregated, the server will distribute information to each participant via different posterior samples of the model's parameters (via the aggregated SS), which are algorithmically calibrated to ensure the derived amount of information is proportionate to the participant's contribution. Such privacy-valuation trade-off will deter participants from selecting excessive DP guarantees that reduce the combined model's utility. Despite this initial success, there remains significant room for improvement. First, this approach is restricted to simple models (e.g., Bayesian linear regression) whose sufficient statistics can be derived analytically and a single round of information sharing, which is a simplified, one-shot communication of federated learning. Second, it has not considered the truthfulness of submitted information and value data as-is. Last, its contribution valuation requires exhaustive enumeration of subsets of participants which does not scale well to large collaborative learning networks. Research in this thrust will focus on addressing these challenges, improving its integrability to large-scale scenarios.

C4. Robust and Sustainable Collaborative Learning: Leveraging the preliminary insights and envisioned contributions from **C1**, **C2** and **C3** (Hoang et al. 2019a; Yurochkin et al. 2019a,b; Hoang et al. 2020; Lam et al. 2021; Bui et al.

2024; Hoang and Hoang 2024), we aim to develop a robust and sustainable solution framework for collaborative learning within an ever-growing ecosystem of learning agents that seek to communicate and reuse relevant knowledge to collaboratively solve their respective tasks. This will complement existing works (Collins et al. 2021; Li et al. 2021; Karimireddy et al. 2020; Hanzely and Richtárik 2020) in FL context which has not considered utilizing such prior knowledge. Furthermore, with the emergence of foundation models which encapsulate vast expertise, a key challenge lies in addressing the sheer scale of these models and seamlessly integrating them into our envisioned resource-aware collaborative learning framework. One important technical aspect that needs to be considered is a potential mismatch between the computational demand of the aforementioned research and the on-board compute capacities of device participants (e.g., drones, wearable devices) in numerous practical scenarios. Another key aspect is the various system constraints across devices (i.e., data collection and model training bandwidth and schedule) which will inevitably create severe data skewness, scarcity and asynchronous information sharing and update. This could lead to or worsen the impact of catastrophic forgetting due to partial accessibility of local knowledge per learning epoch. The dependence on a single coordinating server also represents another critical constraint, which could result in a computational bottleneck, lacking resilience to unstable, erroneous communication. To mitigate this, a decentralized collaborative learning framework is a potential solution, which were previously investigated in a simplified, one-shot federated learning with specific model representation (Hoang et al. 2019b). Part of the envisioned research in this thrust will focus on generalizing the aforementioned research agenda to remove the necessity of the coordinating server, making knowledge communication peer-to-peer and more resilient against erroneous and unstable communication channels.

References

- Bui, L. M.; Huu, T. T.; Dinh, D.; Nguyen, T. M.; and Hoang, T. N. 2024. Revisiting Kernel Attention with Correlated Gaussian Process Representation. In *UAI*.
- Collins, L.; Hassani, H.; Mokhtari, A.; and Shakkottai, S. 2021. Exploiting Shared Representations for Personalized Federated Learning. In *Proc. ICML*, 2089–2099.
- Dhanaraju, M.; Chenniappan, P.; Ramalingam, K.; Pazhanivelan, S.; and Kaliaperumal, R. 2022. Smart Farming: Internet of Things (IoT)-Based Sustainable Agriculture. *Agriculture*, 12(10).
- Fan, Z.; Fang, H.; Wang, X.; Zhou, Z.; Pei, J.; Friedlander, M.; and Zhang, Y. 2024. Fair and Efficient Contribution Valuation for Vertical Federated Learning. In *The Twelfth International Conference on Learning Representations*.
- Ghorbani, A.; Kim, M. P.; and Zou, J. Y. 2020. A Distributional Framework for Data Valuation. *CoRR*, abs/2002.12334.
- Ghorbani, A.; and Zou, J. 2019. Data Shapley: Equitable Valuation of Data for Machine Learning. *arXiv:1904.02868*.
- Hanzely, F.; and Richtárik, P. 2020. Federated Learning of a Mixture of Global and Local Models. *CoRR*, abs/2002.05516.
- Hassantabar, S.; Stefano, N.; Ghanakota, V.; Ferrari, A.; Nicola, G. N.; Bruno, R.; Marino, I. R.; and Jha, N. K. 2020. CovidDeep: SARS-CoV-2/COVID-19 Test Based on Wearable Medical Sensors and Efficient Neural Networks. *CoRR*, abs/2007.10497.
- Hoang, M.; and Hoang, T. N. 2024. Few-Shot Learning via Repurposing Ensemble of Black-Box Models. In *AAAI*.
- Hoang, Q. M.; Hoang, T. N.; Low, K. H.; and Kingsford, C. 2019a. Collective Model Fusion of Multiple Black-Box Experts. In *Proc. ICML*.
- Hoang, T. N.; Hoang, Q. M.; Low, K. H.; and How, J. P. 2019b. Collective Online Learning of Gaussian Processes in Massive Multi-Agent Systems. In *Proc. AAAI*.
- Hoang, T. N.; Lam, C. T.; Low, K. H.; and Jaillet, P. 2020. Learning Task-Agnostic Embedding of Multiple Black-Box Experts for Multi-Task Model Fusion. In *Proc. ICML*.
- Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; and Chen, W. 2021. LoRA: Low-Rank Adaptation of Large Language Models. *CoRR*, abs/2106.09685.
- Itti, L.; and Baldi, P. 2009. Bayesian Surprise Attracts Human Attention. *Vision Research*, 49(10): 1295–1306.
- Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; and Suresh, A. T. 2020. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In *Proc. ICML*, 5132–5143.
- Kingma, D.; and Welling, M. 2013. Auto-Encoding Variational Bayes. In *Proc. ICLR*.
- Konečný, J.; McMahan, H. B.; Ramage, D.; and Richtárik, P. 2016. Federated Optimization: Distributed Machine Learning for On-Device Intelligence. *CoRR*, abs/1610.02527.
- Lam, C. T.; Hoang, T. N.; Low, K. H.; and Jaillet, P. 2021. Model Fusion for Personalized Learning. In *Proc. ICML*.
- Li, T.; Hu, S.; Beirami, A.; and Smith, V. 2021. Ditto: Fair and Robust Federated Learning Through Personalization. In *Proc. ICML*, 6357–6368.
- Li, W.; Fu, S.; Zhang, F.; and Pang, Y. 2023. Data Valuation and Detections in Federated Learning. *ArXiv*, abs/2311.05304.
- McMahan, H. B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proc. AISTATS*, 1273–1282.
- Mironov, I. 2017. Rényi Differential Privacy. In *Proc. 30th IEEE Computer Security Foundations Symposium (CSF)*, 263–275.
- Sim, R. H. L.; Zhang, Y.; Hoang, T. N.; Xu, X.; Low, B. K. H.; and Jaillet, P. 2023. Incentives in Private Collaborative Machine Learning. In *Proceedings of NeurIPS*.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *NeurIPS*, 5998–6008.

Wang, T.; Rausch, J.; Zhang, C.; Jia, R.; and Song, D. 2020. A Principled Approach to Data Valuation for Federated Learning. *CoRR*, abs/2009.06192.

Wang, Z.; Zhang, Z.; Lee, C.; Zhang, H.; Sun, R.; Ren, X.; Su, G.; Perot, V.; Dy, J. G.; and Pfister, T. 2021. Learning to Prompt for Continual Learning. *CoRR*, abs/2112.08654.

Wei, S.; Tong, Y.; Zhou, Z.; and Song, T. 2020. *Efficient and Fair Data Valuation for Horizontal Federated Learning*, 139–152. Cham: Springer International Publishing. ISBN 978-3-030-63076-8.

Yoon, J.; Arik, S. Ö.; and Pfister, T. 2019. Data Valuation using Reinforcement Learning. *CoRR*, abs/1909.11671.

Yurochkin, M.; Agarwal, M.; Ghosh, S.; Greenewald, K.; Hoang, T. N.; and Khazaeni, Y. 2019a. Bayesian Nonparametric Federated Learning of Neural Networks. In *Proc. ICML*.

Yurochkin, M.; Argawal, M.; Ghosh, S.; Greenewald, K.; and Hoang, T. N. 2019b. Statistical Model Aggregation via Parameter Matching. In *Proc. NeurIPS*, 10954–10964.