Federal Register Notice 87 FR 35250, https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies, June 9, 2022

Request for Information (RFI) on

Advancing Privacy Enhancing Technologies

Argonne National Laboratory (ANL)

Response to RFI on Privacy-Enhancing Technologies

Prepared by

Kibaek Kim (Computational Mathematician)
Alec Poczatek (Cybersecurity Analyst)
Dan Harkness (Interim Group Leader)
Minseouk Ryu (Postdoctoral Researcher)
Ravi Madduri (Computer Scientist)

Submitted on July 8, 2022

Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL, 60439

Introduction

About Argonne National Laboratory

Argonne National Laboratory (ANL), one of the U.S. Department of Energy (DOE) national laboratories for science and engineering research, is managed for DOE by UChicago Argonne, LLC. Argonne's mission is to apply a unique mix of world-class science, engineering, and user facilities to deliver innovative research and technologies. Research at Argonne includes energy, biological and environmental systems, advanced computing, and national security. In the following paragraphs, we describe some of the unique capabilities and facilities at Argonne that would help advance the research and development of PETs.

In 2019, Argonne National Laboratory began establishing an AI testbed with following goals: (1) to provide an open and unbiased environment for the evaluation of emerging AI accelerator technologies designed to accelerate training and inference for deep learning models; (2) to document and make available to others information about use cases, software compatibility, software integration requirements, and realized performance on a variety of test problems; and (3) to support collaborations with AI technology developers, academic computer science and data science departments, commercial sector players in AI, and the DOE laboratories. The testbed will work with AI technology companies to make their systems available to the academic, laboratory, and commercial AI developer communities. The specific systems to be deployed in the testbed are not yet known but are expected to include systems such as the Graphcore IPU, Wave Computing DPU, Google Edge TPU, and Loihi neuromorphic chip.

In 2021, Argonne implemented the necessary policies, procedures, controls, and systems to support the analysis of human subject data, called the Argonne Biomedical Learning Environment (ABLE). ABLE is built on Argonne’s moderate enclave that implements NIST 800-53 controls with a separation of duties between administration and research activities. When completed, the enclave will have the necessary privacy controls implemented into all processes providing compliance with HIPAA and FISMA requirements for handling protected health information and personally identifiable information. The ABLE environment at Argonne consists of several high-level resources including cutting-edge hardware for deep learning and machine learning.

In 2022, installation of Aurora, a new Intel-Cray exascale computer, began at the Argonne Leadership Computing Facility (ALCF). When fully deployed, Aurora is projected to have a peak performance of more than 2 exaflops, with over 10,000 nodes and over 10 petabytes of aggregate system memory. The revolutionary architecture will support large-scale simulation (PDEs, traditional HPC), data-intensive applications (scalable science pipelines), and deep learning and emerging science AI (training and inferencing

Argonne Responses

In this section we provide responses to specific topics in the RFI. We use the topic description followed by Argonne's response.

1. Specific research opportunities to advance PETs: Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.

Privacy-preserving federated learning (PPFL)

Federated learning (FL) enables training a machine learning model from distributed data sources without collecting the data to a central location. Specifically, FL is capable of training a model by sharing not the raw data but the processed results (e.g., gradients of the loss function, local model parameters, or hidden representation of the neural network models) between a server and clients (i.e., data owners). This capability is favorable not only to data owners with privacy concerns but also to data analysts who need more data for better learning. It is beneficial especially when the data owners cannot transfer data to the central server because of privacy concerns of citizens or legal frameworks (e.g., medical data) or in areas with national security interests. FL itself, however, cannot guarantee data privacy because the intermediate results communicated during the FL process can be utilized to estimate the local data (e.g., inference attack **[1]**). See Figure 1 generated by the authors for the report to the DOE Advanced Scientific Computing Advisory Committee Meeting in March 2022. This situation thus calls for the development of privacy-preserving FL (PPFL) that integrates privacy-preserving techniques in FL, thereby opening new research directions in the FL community.

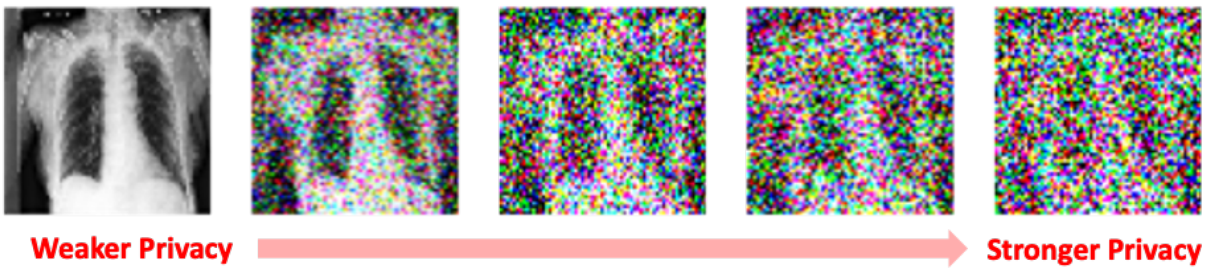

Figure 1. Reconstruction of chest X-ray images from model weights by leveraging the communication during FL with weaker privacy (left) and stronger privacy (right). [23]

PPFL is a recent advance in FL that aims to ensure data privacy by introducing privacy-preservation techniques such as differential privacy, homomorphic encryption, and secure computations. Differential privacy, the state-of-the-art privacy technique for quantifying and limiting information disclosure by random perturbation that is supported by theoretical justifications, has been widely adopted for PPFL mainly because of its computational efficiency compared with other privacy-preserving techniques. For example in **[12],** the computational efficiency of differential privacy is compared with homomorphic encryption. Differentially private FL, however, encounters inevitable trade-offs between data privacy and learning performance (i.e., more noise added for ensuring stronger data privacy downgrades the learning performance). Needed, therefore, are novel PPFL algorithms that provide higher accuracy under strong data privacy. For example, the inexact alternating direction method of multipliers algorithm developed in our recent work **[17]** provides the same level of differential privacy as the state-of-the-art PPFL algorithm but outperforms it by exploiting the objective perturbation for differential privacy and multiple local update techniques. *Among the research directions in PPFL, critical are the development of advanced training algorithms and communications with differential privacy and other privacy-preserving techniques.*

More challenges and opportunities exist and arise from different settings of FL, each of which further complicates privacy-preserving procedures. First, data distributed across the clients can be unbalanced and not identically and independently distributed (IID). Non-IID data can make the model training challenging. For example, in the cross-device setting (e.g., millions of sensors and edge devices), a subset of clients may be sampled at every iteration of a global update, resulting in the underlying distributions changing at every step of the algorithm. Another challenge is the communication required for training a model across the clients. On the one hand, FL with data distributed in a few large silos (i.e., cross-silo FL) may experience computational load imbalance across the clients, causing some local model trains to take more time than others. On the other hand, in the case of cross-device FL, a large number of devices may suffer from communication bottlenecks with limited network capacities. Moreover, in both cross-silo and cross-device settings, some clients may be temporarily unavailable or fail during the training. Efficient communication between a server and clients is also a key challenge in PPFL as more communication rounds could increase the risk of data leakage because more intermediate results can be utilized for the reverse-engineering process for estimating the locally stored data. Therefore, advanced communication strategy (e.g., asynchronous update, network topology, and compression) needs to be explored.

PPFL on multimodal datasets

Data modality will also lead to different settings of FL. Depending on the distributed data modality, FL can also be categorized as (i) horizontal FL (HFL) **[2,3,4,5,6,7]** or (ii) vertical FL (VFL) **[8,9,10,11,12]**. In the HFL setting, every client shares the same data features but different data samples, which can be considered as a data matrix (i.e., data samples and features are

represented by rows and columns, respectively) split horizontally and distributed over multiple clients. In the VFL setting, every client shares the same data samples but different data features, which can be considered as the data matrix split vertically and distributed over multiple clients. The key difference between HFL and VFL is that HFL is limited to unimodal data, whereas VFL is applicable to multimodal data. We observe that VFL has been underexplored, whereas HFL has been actively studied in both academics and practice. Moreover, from the privacy-preserving perspective, we expect that achieving data privacy will be more challenging with VFL than with HFL, mainly because of the expensive and complicated communication required for multimodal data in VFL. Moreover, most FL has been developed for supervised learning. In many applications (e.g., anomaly detection using satellite data **[13,14]** and feature training from multimodal data **[15,16]**), data labels are often missing, calling for developing FL algorithms for unsupervised HFL and VFL.

Trusted AI using PETs

Trusted AI models are resilient to model shift and preserve the privacy of data used in training. To enable the creation of trusted AI in PETs, advances are needed in building trusted data repositories and cyber infrastructure practices that enable continuous training of models resilient to bias and ensure fairness to improve the overall confidence of the general public. Achieving these advances requires the development of well-labeled (also known as AI-ready) data, managed by using FAIR (Findability, Accessibility, Interoperability, and Reproducibility) principles along with well-documented APIs to securely access the data. Publicly available AI-ready datasets are important in developing baseline models that can later be used in implementing continuous training and validation of AI models. These, coupled with the implementation of a “trusted model repository,” can provide AI practitioners details on provenance and training data distributions, along with model parameters that are useful in assessing the quality and maturity of the model.

Data and model shifts in AI models dent the applicability of AI in real-world settings. Models are better performing when they are trained from the real world and are continuously updated to accurately capture changes in data distribution. However, real-world data in many domains is often hard to access because of privacy concerns and reidentification attacks. PPFL offers a framework to address the data and model shifts in AI models, especially when trained on real-world data with sufficient privacy guarantees. ***We envision the creation of a continuously updated, trusted AI model repository trained by using disparate private data, leveraging PPFL technologies to mitigate bias and improve confidence in the applicability of AI models to the real world***.

Continuous training and validation of AI models using PPFL technologies require a fair amount of computational power depending on the privacy-budget allocated to the differential privacy algorithm. Stronger privacy protection means a larger privacy budget and hence more computational power. **We envision leveraging DOE’s supercomputing facilities to provide the computing required to create and maintain the trusted model repository**. Frameworks such as Parsl **[24],** FuncX **[25**], and Flux **[26**] that support for distributed computational

workloads exist abut need to be integrated with PPFL frameworks to fulfill the vision. PPFL frameworks should be integrated in IAM (identity and access management) systems to leverage secure data transfer technologies and ensure that large-sized models and datasets can be transferred reliably.

DOE runs the world's fastest supercomputers and more recently has set up private enclaves that are certified to host private data securely. Leveraging PETs to build trustworthy AI models that are robust and up to date requires a comprehensive redesign of data management and computational strategies to create an ecosystem of capabilities. **We envision the creation of a comprehensive framework that evaluates models trained using PPFL technologies against various threats and attack models.**

Confidential computing on leadership-class systems

Confidential computing uses hardware-based techniques to isolate data, specific functions, or an entire application from the operating system, hypervisor or virtual machine manager, and other privileged processes. Data is stored in the trusted execution environment (TEE), where it is impossible to view the data or operations performed on it from outside, even with a debugger. The TEE ensures that only authorized code can access the data. If the code is altered or tampered with, the TEE denies the operation. Technologies such as HPCrypt **[27]** allow creation of secure private enclaves on supercomputing resources using cryptographic technologies that are integrated into policies that govern execution on supercomputers. To enable confidential computing on leadership-class computing facilities**, we envision integration of differential privacy with privacy-enhancing cryptography techniques** such as secure multiparty computation, private set intersection, private information retrieval, zero-knowledge proofs, and fully homomorphic encryption with TEEs.

2. Specific technical aspects or limitations of PETs: Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

Evaluation of existing PPFL packages

A few open-source FL frameworks exist. These include Argonne Privacy-Preserving Federated Learning (APPFL) **[18]**, Open Federated Learning (OpenFL) **[19]**, Federated Machine Learning (FedML) **[20]**, TensorFlow Federated (TFF) **[21]**, and PySyft **[22]** The existing software packages have been developed to implement various FL algorithms with different

communication protocols while leveraging the existing ML tools such as PyTorch and TensorFlow. Some packages also implement privacy-preserving techniques such as differential privacy. Here we briefly discuss the capabilities of each framework.

APPFL: This package has been developed by Argonne National Laboratory via multiple DOE-ASCR projects. It provides PPFL tools for users in practice while allowing research communities to implement, test, and validate various ideas for PPFL. The current implementation includes a number of training algorithms with differential privacy. The package supports the PyTorch data and model objects and the communication via MPI or gRPC. The package was highlighted at the DOE Advanced Scientific Computing Advisory Committee Meeting in March 2022.

OpenFL: This framework has been developed by Intel. It was initially developed as part of a research project on FL for healthcare and designed for a multi-institutional setting. In OpenFL, an environment is constructed based on collaborator and aggregator nodes that form a star topology; in other words, all collaborator nodes are connected to an aggregator node. Communication between nodes is through gRPC via a mutually authenticated transport layer security network connection. However, OpenFL does not support different communication protocols other than gRPC which limits adoption in heterogenous computing environments.

FedML: This is an open research library to facilitate FL algorithm development and fair performance comparison. It supports on-device training for edge devices, distributed computing, and single-machine simulation. It utilizes gRPC and MQTT for device communication to simulate cross-device FL on real-world hardware platforms. Also, it utilizes MPI for simulating FL in a distributed-computing setting. It implements weak differential privacy that aims to prevent a backdoor attack, which requires less noise in training data compared with what is required for ensuring data privacy. However, FedML does not support adding custom differential privacy algorithms that limits it applicability to track and adopt the latest developments in the space.

TFF: This framework is available from Google for machine learning and other computations on decentralized data. In TFF, an FL environment is constructed by using multiple GPUs that are used as clients. Also, TFF can be simulated on a Google Cloud platform. Currently, TFF supports FedAvg and differential privacy for private federated learning. TFF only supports gRPC-based communication protocols that limit the applicability in heterogenous computing environments.

PySyft: This framework is available from OpenMined, an open-source community. In PySyft, an FL environment is constructed by Virtual Workers, WebSocket Workers, or GridNodes. While Virtual Workers live on the same machine and do not communicate over the network, the others leverage WebSocket as a communication medium to ensure that a broad range of devices can participate in a PySyft network. Currently, PySyft supports FedAvg and differential privacy for private federated learning.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs:Information about sectors, applications, or types of analysis that have high potential for the adoption of PETs. This includes sectors and applications where data are exceptionally decentralized or sensitive, where PETs could unlock insights or services of significant value to the public, where PETs can reduce the risk of unintentional disclosures, where PETs might assist in data portability and interoperability, and sectors and applications where the adoption of PETs might exacerbate risks, including in the areas of privacy, cybersecurity, accuracy of data analysis, equity for underserved communities, and economic competition. This topic covers opportunities to improve the effectiveness of data sharing among specific Federal agencies and between specific Federal agencies and entities outside the Federal Government, including the goals outlined in Section 5 of Executive Order 14058: Transforming Federal Customer Experience and Service Delivery To Rebuild Trust in Government.

Sectors and application domains

Biomedical applications. In biomedicine, data fusion and centralized AI analyses are difficult to realize because of privacy concerns and reidentification attacks. This situation has resulted in multimodal biomedical datasets governed by federal regulations and consortium-specific data usage agreements that have made centralized data collection and analysis difficult. PPFL provides a great opportunity to unlock biomedical insights from multimodal data. Applications include building better epidemiological models by combining multimodal private data from different healthcare providers across the international borders, for example for early detection of pandemics and better risk prediction models for various diseases across different populations. The applications of PETs, especially PPFL, have the potential to create robust AI models **[28]**.

DOE user facilities. DOE's Basic Energy Sciences (BES) program has identified its first priority research opportunity as follows: "Efficiently extract critical and strategic information from large, complex datasets at BES's scientific user facilities." User facilities such as Argonne's Advanced Photon Source (APS) are valuable resources that generate multimodal data for numerous different types of use cases. Some of these cases involve Controlled Unclassified Information acquired in support of projects related to NNSA, DOD, and other agencies. However, sharing raw data across geographically separated instruments, especially across institutions, is challenging; and there are concerns (e.g., citizen privacy, legal frameworks for medical data) that those experiments could not make full use of advances in AI/ML or in areas with national security interests. If raw data can stay in place by sharing only the processed results, it can make more information beneficial to larger communities without explicitly sharing any sensitive data.

Critical Infrastructure. The electric power grid is at the heart of a complex system of interdependent critical infrastructures. It represents many of the common characteristics and challenges of complex engineering systems. Any disruption to the system can lead to dire consequences for the functioning of society and the economy. The electric power grid is undergoing a modernization currently pushed by 100% clean energy targets and enhanced connectivity through distributed sensing and control devices (e.g., PMUs, smart meters). Future

grid operation is anticipated to be distributed/decentralized, as opposed to the existing centralized operations, by utilizing a large amount of distributed data available at the edges of the network. However, such data collected by edge devices may not be available to the central operator in real time because of the large volume of the datasets and data privacy. At the same time, the electric grid is facing complex operational challenges due to increasing threats from natural disasters (e.g., hurricanes, wildfires) and human-initiated cyber incidents. In this context, an important challenge for complex engineering systems is data-driven decision making for resilient operation by accounting for the distributed/decentralized monitoring and control. While potentially reducing the impacts of single-point failures, distributed or decentralized operations introduce new challenges associated with secure data integration from different sources. These challenges need to be addressed by an AI-driven/assisted monitoring and control system with the overarching constraints of data privacy and security.

Cybersecurity federated models. Cybersecurity is often seen as a single-organization-centric model. However, adversaries often target multiple organizations using the same tactics, techniques, and procedures. Cybersecurity federated models (e.g., Argonne's CFM) are community-based security models to promote a global defense against common threats within large heterogeneous and distributed organizations. Many of these models have already approached privacy-enhancement through anonymization and obfuscation (removing identifying information) or through encryption (limiting who can view the information). Anonymization and obfuscation techniques tend to overscrub information to minimize the risk of missing identifiable elements. Encryption does not actually fix the problem but instead just limits the exposure to those trusted to decrypt it. Cybersecurity federated models hold much promise for moving beyond simple information sharing to instead empowering collaborative analysis. Such capabilities, however, will require better privacy-enhancing techniques. With the right assurances of privacy, the future of cybersecurity federated models could include distributed analytics at the edge; distributed query and response capabilities, allowing trusted analysts to locate relevant data across their peers; and federated learning to improve artificial intelligence for cybersecurity.

Cybersecurity improvements with user-behavior. A significant effort within the U.S. government to push toward tighter security models such as zero trust. This direction requires that security architectures have sufficient data to determine whether any action is a malicious one in order to prevent exploits from occurring in near-real time. Preservation of privacy is imperative in this regard, making data difficult to obtain as de-anonymization becomes easier with more available data points to analyze.

Many cybersecurity challenges also stem from the need to be able to de-anonymize data in the event of legal requirements where information must be used in investigations or other scenarios in which the digital information is relevant. This data could be the only evidence of data breaches, data manipulation, or other malicious activity, making privacy protections that much harder to balance.

Data collection such as packet captures of internet traffic in an organization also stand to be a valuable resource in detecting malicious behavior. But because such packet captures contain all

data sent by a given user as well as the data received by that user, they pose a significant potential for privacy issues. Indeed, the captured data can include passwords, personally identifiable information, email contents, schedules, and myriad other personal data. It also may include company intellectual property and any other type of data that gets sent across a network. This type of data is one of the hardest types of datasets for cybersecurity researchers to obtain, since very few datasets have been curated enough to ensure that none of this privacy infringing information is contained within it.

Fingerprinting, or tracking users based upon correlating factors on how they interact with computer systems, is also a common technique utilized by cybersecurity operations teams to identify whether the user is a known threat. Implementing user fingerprinting is utilized heavily in advertising agencies as well to try to sell people products as more information is gathered about their personal interests and spending habits. The same technology utilized by these advertising agencies is used to determine whether users are attempting to access data that they shouldn't be. The aggregated data gathered about these users can allow for the identification of a specific person. Here PETs can assist with ensuring that user privacy ise maintained while still ensuring that threats are identified.

6. *Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs:* This includes the development of open-source protocols and technical guidance, the use of public-private partnerships, prize challenges, grants, testbeds, standards, collaborations with foreign countries and nongovernmental entities, the Federal Data Strategy, and data sharing procedures with State, local, tribal, and territorial governments. This also includes interpretations and modifications of standard non-disclosure agreements, confidentiality clauses, data use or sharing agreements, etc.

The Privacy Group under the Cybersecurity Priority area of the National Institute for Standards and Technology has developed comprehensive guidelines and evaluated tools developed by academia and leading internet companies that can be incorporated in various application domains.

References

[1] Shokri, Reza, et al. "Membership inference attacks against machine learning models." *2017 IEEE symposium on security and privacy (SP)*. IEEE, 2017.
[2] Konečný, Jakub, et al. "Federated learning: Strategies for improving communication efficiency." *arXiv preprint arXiv:1610.05492* (2016).
[3] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." *Artificial intelligence and statistics*. PMLR, 2017.
[4] Wei, Kang, et al. "Federated learning with differential privacy: Algorithms and performance analysis." *IEEE Transactions on Information Forensics and Security* 15 (2020): 3454-3469.
[5] Truex, Stacey, et al. "LDP-Fed: Federated learning with local differential privacy." *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*. 2020.
[6] Huang, Zonghao, et al. "DP-ADMM: ADMM-based distributed learning with differential privacy." *IEEE Transactions on Information Forensics and Security* 15 (2019): 1002-1012.
[7] Agarwal, Naman, et al. "cpSGD: Communication-efficient and differentially-private distributed SGD." *Advances in Neural Information Processing Systems* 31 (2018).
[8] Wei, Kang, et al. "Vertical Federated Learning: Challenges, Methodologies and Experiments." *arXiv preprint arXiv:2202.04309* (2022).
[9] Chen, Tianyi, et al. "Vafl: a method of vertical asynchronous federated learning." *arXiv preprint arXiv:2007.06081* (2020).
[10] Romanini, Daniele, et al. "PyVertical: A vertical federated learning framework for multi-headed SplitNN." *arXiv preprint arXiv:2104.00489* (2021).
[11] Sun, Jiankai, et al. "Vertical federated learning without revealing intersection membership." *arXiv preprint arXiv:2106.05508* (2021).
[12] Wang, Chang, et al. "Hybrid differentially private federated learning on vertically partitioned data." *arXiv preprint arXiv:2009.02763* (2020).
[13] Liu, Yi, et al. "Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach." *IEEE Internet of Things Journal* 8.8 (2020): 6348-6358.
[14] Zhang, Kai, et al. "Federated variational learning for anomaly detection in multivariate time series." *2021 IEEE International Performance, Computing, and Communications Conference (IPCCC)*. IEEE, 2021.
[15] Ngiam, Jiquan, et al. "Multimodal deep learning." *ICML*. 2011.
[16] Zhao, Yuchen, Payam Barnaghi, and Hamed Haddadi. "Multimodal Federated Learning on IoT Data." *2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 2022.
[17] Ryu, Minseok, and Kibaek Kim. "Differentially private federated learning via inexact ADMM with multiple local updates." *arXiv preprint arXiv:2202.09409* (2022).
[18] Ryu, Minseok, et al. "APPFL: Open-Source Software Framework for Privacy-Preserving Federated Learning." *In: Proceedings of the 36th IPDPS 2022 Workshops (to appear)*. 2022.
[19] Reina, G. Anthony, et al. "OpenFL: An open-source framework for Federated Learning." *arXiv preprint arXiv:2105.06413*(2021).

[20] He, Chaoyang, et al. "FedML: A research library and benchmark for federated machine learning." *arXiv preprint arXiv:2007.13518* (2020).
[21] Bonawitz, K., H. Eichner, and W. Grieskamp. "TensorFlow federated: machine learning on decentralized data.(2020)."
[22] Ziller, Alexander, et al. "Pysyft: A library for easy federated learning." *Federated Learning Systems*. Springer, Cham, 2021. 111-139.
[23] Helland, Barbara. "View from Germantown: Advanced Scientific Computing Research Update". U.S. Department of Energy, Office of Science. https://science.osti.gov/-/media/ascr/ascac/pdf/meetings/202203/ASCAC_202203_Presentation_Helland.pdf Accessed: 2022-07-07
[24] Y. Babuji, et al. 2019. Parsl: Pervasive Parallel Programming in Python. In 28th International Symposium on High-Performance Parallel and Distributed Computing (HPDC'19). ACM, 25--36.
[25] Ryan Chard, et al. 2020. FuncX: A Federated Function Serving Fabric for Science. HPDC '20. Association for Computing Machinery, New York, NY, USA, 65–76.
[26] Dong H. Ahn, et al, “Flux: Overcoming Scheduling Challenges for Exascale Workflows”, Future Generation Computer Systems, Volume 110, 2020, Pages 202-213
[27] HPCrypt Data Protection System,developed by the Lawrence Livermore National Laboratory, https://ipo.llnl.gov/technologies/hpcrypt
[28] The clinician and dataset shift in artificial intelligence. The New England Journal of Medicine, 2021