PhD-FSTM-2023-082
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 26/09/2023 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN INFORMATIQUE

by

Joaquín DELGADO FERNÁNDEZ
Born on 8 February 1996 in Salamanca, (Spain)

BREAKING DATA SILOS WITH FEDERATED LEARNING

Dissertation defence committee

Dr Gilbert Fridgen, dissertation supervisor
Professor, Université du Luxembourg

Dr Gerhard Schwabe
Professor, University of Zurich

Dr Radu State, Chairman
Professor, Université du Luxembourg

Dr Robert Keller
Professor, Kempten University of Applied Sciences

Dr Miguel Angel Olivares Mendez, Vice-Chairman
Assistant Professor, Université du Luxembourg

“Quise ser valiente y aprendí a estudiar.”

— Antonio Escohotado Espinosa (1941 – 2021).

“- Tú dices que la inspiración no existe.”

“- No existe, no... la inspiración no... no soy yo el que dice que no existe. Era Baudelaire. Cuando le preguntó una señora: "¿Qué es la inspiración, maestro?". Le contestó: "Señora, la inspiración es trabajar todos los días". Claro, yo me siento a la mesa de escribir y la inspiración acaba llegando.”

— Camilo José Cela, *A Fondo (1989)*.

Acknowledgements

Primeramente, a mis padres y a mis abuelos. Esta tesis no habría sido posible sin su inestimable ayuda y sin sus palabras de ánimo. Por esos *tranquilo, seguro que todo irá bien*. Gracias.

À ma copine Anna, qui est à la fois le phare de l'incertitude et le gouvernail de mon ambition. Quand tout change, quand tout vacille, elle est là. Merci.

A mis amigos, Alberto, Miguel y Mario, *los de toda la vida*, que desde la distancia, siempre disponibles y atentos. Siempre con esa dura pregunta *¿Cuándo vuelves?* en las despedidas. Gracias.

To my friends and colleagues at Finatrax: Sergio, Tom, Esti, Iván, Linda, Orestis, Reilly, Charles, Renan, Timothée and many more. You made the Ph.D. a little less lonely and much more enjoyable. This journey would have been much harder without you. Thanks, you will be missed.

Lastly, to my supervisor, Gilbert and my *de-facto* mentor Alex, I am grateful for your invaluable advice, continuous help, exciting ideas, and patience during my Ph.D. study.

Abstract

Federated learning has been recognized as a promising technology with the potential to revolutionize the field of Artificial Intelligence (AI). By leveraging its decentralized nature, it has the potential to overcome known barriers to AI, such as data acquisition and privacy, paving the way for unprecedented advances in AI.

This dissertation argues the benefits of this technology as a catalyst for the irruption of AI both in the public and private sector. Federated learning promotes cooperation among otherwise competitive entities by enabling cooperative efforts to achieve a common goal.

In this dissertation, I investigate the goodness-of-fit of this technology in several contexts, with a focus on its application in power systems, financial institutions, and public administrations. The dissertation comprises five papers that investigate various aspects of federated learning in the aforementioned contexts. In particular, the first two papers explore promising venues in the energy sector, where federated learning offers a compelling solution to privately exploit the vast amounts of data and decentralized ownership of data by consumers. The third paper elaborates on another paradigmatic example, in which federated learning is used to foster cooperation among financial institutions to produce accurate credit risk models. The fourth paper makes a juxtaposition with the previous ones centered on the private sector. It elaborates on the use cases of federated learning for public administrations to reduce barriers to cooperation. Lastly, the fifth and last article acts as a *finale* of this dissertation, compiles the earlier work and elaborates on the constraints and opportunities associated with adopting this technology, as well as a framework for doing so.

Resumé

L'apprentissage fédéré a été reconnu comme une technologie prometteuse ayant le potentiel de révolutionner le domaine de l'intelligence artificielle (IA). En exploitant sa nature décentralisée, il a le potentiel de surmonter les obstacles connus en matière d'IA tels que l'acquisition de données et la confidentialité, ouvrant la voie à des avancées sans précédent dans le domaine de l'IA.

Cette thèse défend les avantages de cette technologie en tant que catalyseur de l'irruption de l'IA tant dans le secteur public que privé. L'apprentissage fédéré favorise la coopération entre des entités autrement concurrentes en permettant des efforts coopératifs pour atteindre un objectif commun.

La pertinence de cette technologie est étudiée dans plusieurs contextes, en mettant l'accent sur son application dans les systèmes électriques, les institutions financières et les administrations publiques. La thèse comprend cinc articles qui examinent différents aspects de l'apprentissage fédéré dans les contextes susmentionnés. En particulier, les deux premiers articles explorent des perspectives prometteuses dans le secteur de l'énergie, où l'apprentissage fédéré offre une solution intéressante pour utiliser les vastes quantités de données et la propriété décentralisée des données par les consommateurs de manière privée. Le troisième article développe un autre exemple paradigmatique, dans lequel l'apprentissage fédéré est utilisé pour favoriser la collaboration entre les institutions financières afin de produire des modèles précis de risque de crédit. Le quatrième article établit une juxtaposition avec les précédents, centrés sur le secteur privé. Il développe les cas d'utilisation de l'apprentissage fédéré pour les administrations publiques afin de réduire les obstacles à la collaboration. Enfin, le cinquième et ultime article permet de conclure cette thèse, compilant les travaux antérieurs et développant les contraintes et opportunités associées à l'adoption de cette technologie, ainsi qu'un cadre pour le faire.

Resumen

El aprendizaje federado es reconocido como una tecnología prometedora con el potencial de revolucionar el campo de la Inteligencia Artificial (IA). Al aprovechar su naturaleza descentralizada, tiene la capacidad de superar barreras conocidas en IA como son la adquisición de datos y la privacidad, abriendo camino para avances sin precedentes.

Esta tesis argumenta los beneficios de esta tecnología como catalizador para la irrupción de la IA tanto en el sector público como en el privado. El aprendizaje federado promueve la cooperación entre entidades que de otra manera serían competidoras, al permitir esfuerzos cooperativos para alcanzar un objetivo común.

La idoneidad de esta tecnología se investiga en varios contextos, con un enfoque en su aplicación en el sector eléctrico, instituciones financieras y administraciones públicas. La tesis consta de cinco artículos que investigan diversos aspectos del aprendizaje federado en los contextos mencionados. En particular, los primeros dos artículos exploran oportunidades prometedoras en el sector energético, donde el aprendizaje federado ofrece una solución optimista para utilizar la gran cantidad de datos y la propiedad descentralizada de los mismos por parte de los consumidores de manera privada. El tercer artículo profundiza en otro ejemplo paradigmático, en el que el aprendizaje federado se utiliza para fomentar la colaboración entre instituciones financieras y producir modelos precisos de riesgo crediticio. El cuarto artículo realiza una yuxtaposición con los anteriores centrados en el sector privado. Elabora sobre los casos de uso del aprendizaje federado para las administraciones públicas con el objetivo de reducir barreras a la colaboración. Por último, el quinto artículo actúa como una *finale* de esta tesis, recopila los trabajos anteriores y elabora sobre las limitaciones y oportunidades asociadas con la adopción de esta tecnología, así como un marco para hacerlo.

Table of Contents

List of Figures

I Introduction

The exponential growth of connectivity between humans has generated an unprecedented amount of raw data. This explosion of data provides significant potential for businesses to derive valuable insights that can drive innovation and progress. Machine learning (ML) techniques have appeared as a crucial tool to unlock the potential of data. ML and by extension Deep Learning (DL) [1] can autonomously learn from large datasets and find complex relationships and patterns, allowing businesses to make data-driven decisions (McAfee et al., 2012).

The increasing value of data and its resulting scarcity characterize the modern data landscape (Attard et al., 2016). Data has become a crucial source of competitive advantage for companies, allowing them to better estimate the future of their products or operations (Manyika et al., 2011).

In practice, data-intensive companies can form collaborative alliances to increase their access to data. These alliances typically resort to using data lakes (Fang, 2015; Giebler et al., 2019) - shared pools of data where entities can participate and use each other's data. Although data lakes can address fundamental challenges, such as data availability or reducing management costs, they also raise several concerns, particularly about the privacy of shared data (Chen et al., 2012; Eder and Shekhovtsov, 2020). This creates a challenging situation where entities can benefit from the existence of a data lake, but must invest significant resources to anonymize and ensure the quality of shared data (Eurich et al., 2010). In addition, entities may choose to selectively disclose low-quality data to reduce the probability of losing a competitive edge.

[1] DL is a subfield of ML that focuses on the training of multilayer neural networks. These (deeper) networks aim to extract meaningful patterns and representations from complex data.

"*Unfortunately, the privacy issues due to data sharing remains a missing piece in the data lake designs*" (Chen et al., 2018). Privacy issues, particularly those related to the existence of highly sensitive and private personal data, can delay the adoption of data lakes. To overcome these limitations, McMahan et al. (2017) proposed federated learning (FL). FL is a distributed machine learning paradigm that enables multiple decentralized clients to collaboratively train a global model. It utilizes each client's own local training dataset, without compromising the privacy of the underlying data nor accessing it. FL is coordinated by a central server that acts as a trusted intermediary to prevent malicious or collusive behavior, according to the general Byzantine problem (Lamport et al., 1982). Notably, the client's original datasets remain confidential and inaccessible to both the coordinator and other clients.

This thesis investigates the potential of FL in three distinct sectors, namely energy, finance, and public. It examines the advantages and difficulties of applying FL in these domains and emphasizes the positive impact of FL on overcoming the inherent data constraints in these sectors. Firstly, this dissertation explores the effects of FL for short-term load forecasting (STLF) in the energy sector, it utilizes peer-to-peer (P2P) clustering to reduce variability between households, and develops privacy preserving techniques such as Differential Privacy (DP) and secure aggregation (SecAgg) to protect individuals. Secondly, this dissertation deepens the potential of FL in the financial sector by investigating the feasibility of creating distributed credit risk models between financial institutions. It explores the possibility of small financial institutions collaborating to develop highly efficient and precise credit risk models. Thirdly and lastly, this dissertation investigates the public sector's applicability of FL to address the challenges posed by stringent regulations and improve communication between governments.

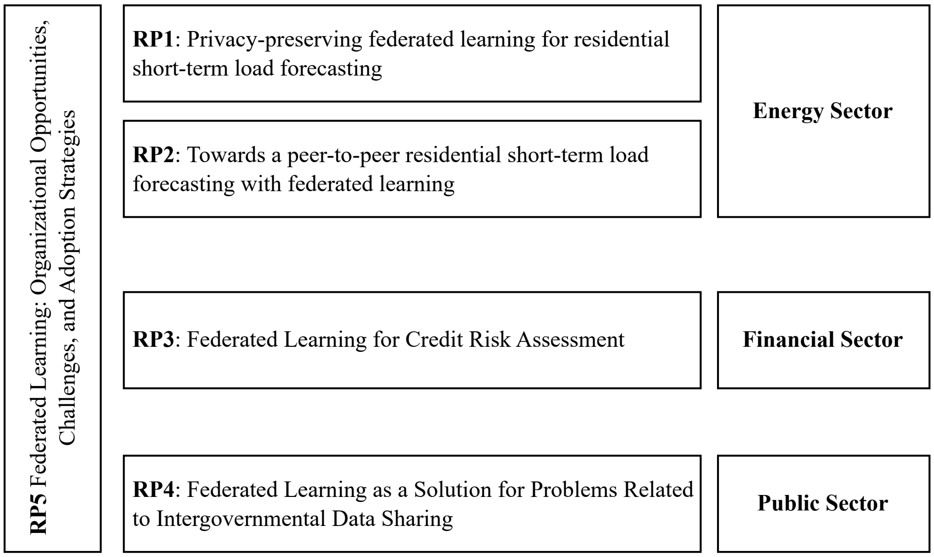

RP5 Federated Learning: Organizational Opportunities, Challenges, and Adoption Strategies
RP1: Privacy-preserving federated learning for residential short-term load forecasting
RP2: Towards a peer-to-peer residential short-term load forecasting with federated learning
Energy Sector
RP3: Federated Learning for Credit Risk Assessment
Financial Sector
RP4: Federated Learning as a Solution for Problems Related to Intergovernmental Data Sharing
Public Sector

Figure I.1: Structure of the dissertation

This cumulative thesis is compromised by five publications and is structured as follows (see Figure I.1). Firstly, two publications related to the applications of FL in electric power systems. That is, RP 1: “*Privacy-preserving federated learning for residential short-term load forecasting* ” where we investigate the effects of DP and FL for STLF. The second publication in power systems: RP 2: “*Towards a peer-to-peer residential short-term load forecasting with federated learning*” explores the possibility of creating a fully decentralized forecasting system. The third publication; RP 3: “*Federated Learning for credit risk assessment*” relates to a possible FL model to overcome data limitations in the financial sector. Later, the fourth publication: RP 4: “*Federated Learning as a Solution for Problems Related to Intergovernmental Data Sharing*” explores how the government should use FL to streamline communication issues. Lastly, a final publication RP 5: “*Federated Learning: Organizational Opportunities, Challenges, and Adoption Strategies*” that builds on previous publications and discusses the implications of FL in information systems research, as well as the market conditions for the adoption of FL projects.

II Foundations of Federated Learning

1 Characterizing Federated Learning

The quality and quantity of data are the main drivers of the effectiveness of AI-based methods. The gathering of large amounts of data in a single silo is often limited due principally to competition, strict regulations on data, privacy and security (Kairouz et al., 2021).

In an era where data volumes have reached unprecedented heights, new approaches are required to effectively use and handle them. FL was originally proposed as a method "*to train a high quality centralized model while training data remains distributed over a large number of clients, each with unreliable and relatively slow network connections*" (Konečný et al., 2016).

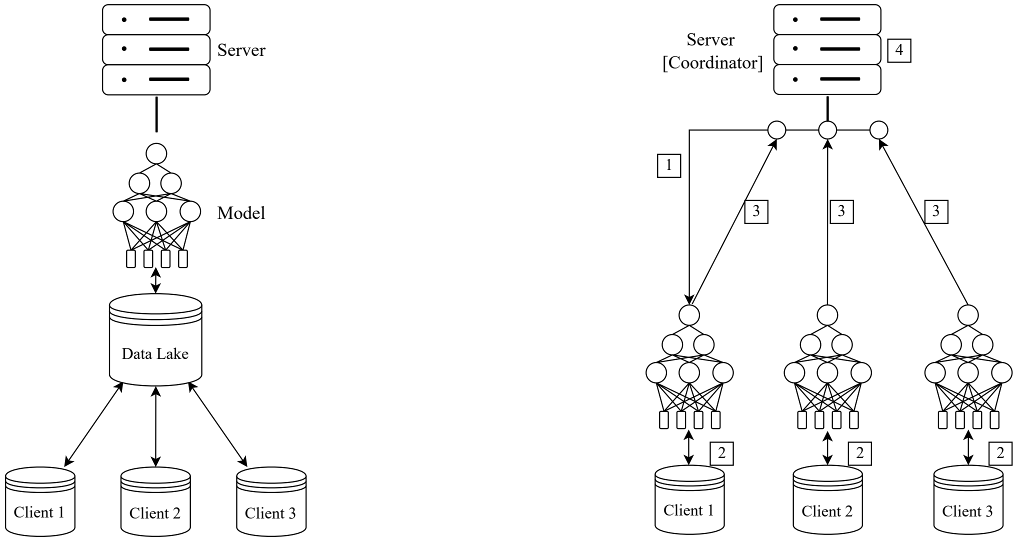
FL aims to break down data silos by providing an efficient solution for leveraging large amounts of decentralized data by facilitating collaborative learning without the need to store data in a central location. In turn, this decentralized approach serves to break down the barriers to data sharing that may exist between different entities. It enables knowledge sharing and collaboration without sharing raw data. Since the train resides on client devices, FL enables organizations to leverage each other's data without compromising privacy or violating regulatory requirements.

1.1. Centralized and Federated Learning comparison

In traditional centralized learning architectures, the training process is twofold. On the one hand, there are cases where the data and the means of processing it are in the same place. In these situations, the data is kept within their boundaries in preparation for further injection into the models during training. On the other hand, where the data is decentralized but the ownership of the exploitation means is centralized, there is an explicit need to anonymize the data to then share it with any data aggregator such as, for instance, the so-called *data lake* that will store them until a centralized model trains on it.

The training process in FL has fundamental differences from traditional machine learning. In FL, data is decentralized in the so-called clients. Each client has sole ownership of its data and thereby is responsible for training its own model. Training of FL models can be seen as a higher-level abstraction of traditional machine learning training. Both methods involve training in rounds or epochs, but in FL, the convergence direction is determined not only by the data of each client, but also by the directions of other clients. This involves a dual process of local training and collective aggregation that continues until a desired level of performance is reached.

Figure 1 presents the differences of the two different architectures, centralized and FL. On the left side (1a) the classic training of a centralized machine learning model is depicted as such. First, the clients will share the anonymized data with the data lake or data aggregator. The data aggregator in step number two will store the data serving as a silo for the model to train in the third step. In the last step, after the centralized model has converged, the server will transmit the model trained back to the initial clients when a predefined performance is achieved. On the right side (1b), the FL process is depicted. Initially, the central server shares a predefined model with the clients. This step ensures the same baseline for model training. Second, clients train their own model based on their own data and share it with the central server in step number three. Fourth, and last, the central server will aggregate the models (usually by averaging them). This process of training, aggregating and sharing repeats until a certain level of performance is achieved.



(a) Centralized.

(b) Decentralized/Federated.

Figure II.1: Architectures for machine learning. (c.f. (Fernández et al., 2023a RP 5))

1.2. Canonical Federated Learning algorithms

There are two canonical training algorithms for FL. Both are composed of the same idea: no raw data is transmitted to any client nor the central server. Instead, the models are trained locally, shared with the central server, aggregated, and then rerouted to the clients.

Originally proposed by McMahan et al. (2017), in Federated Stochastic gradient descent (Fed-SGD), clients send the gradient (*direction*) of the loss to be followed. The direction is generated locally on the clients' own data. Then, it is sent to the server for averaging[1] The averaged direction is finally shared back to the clients, who will apply it to move their own local model towards a lower error rate.

Contrary, in Federated Average (Fed-Avg) clients instead of transmitting the gradients at every batch they process, they transmit an update of the model to the central server after each round of the learning algorithm (McMahan et al., 2017).

The relative superiority of Fed-Avg over Fed-SGD is related to accuracy and the number of communication rounds. Fed-Avg is generally more performant and resilient to disparities in models weights. Moreover, since the model weights are shared and not the

[1] Typically, simple or weighted averages are used, but a large body of research is exploring different averaging techniques. See (Sah and Singh, 2022).

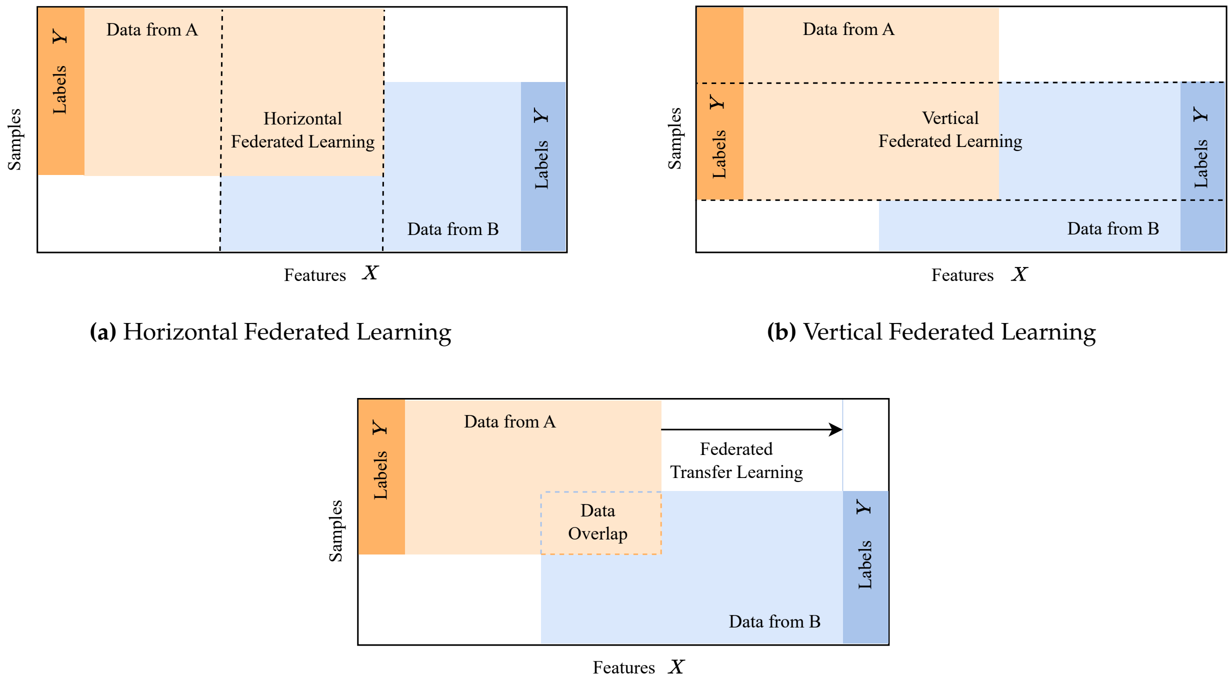
gradient at the end of every batch, communication rounds needed between the central server and the federated parties is reduced (Fekri et al., 2021; McMahan et al., 2017). The relatively lower number of communication rounds has positive effects on bandwidth, the stress put on processing units, and the time required to train a model. As a consequence, it may also reduce energy consumption. Despite the differences mentioned above, both Fed-SGD and Fed-Avg are an iterative process to gradually reduce the prediction error compared to the ground truth. During these steps, the models take steps towards a lower error rate.

While the main canonical implementations rely on a central server, new advances have pave the way for a fully decentralized FL. While this thesis is built upon the original work of (McMahan et al., 2017) and (McMahan et al., 2018), I encourage the interested readers to become familiar with serverless implementations of FL (Chang et al., 2018; Kalra et al., 2023; Shen et al., 2020)

1.3. Federated Learning settings

In the same way there are different training algorithms for FL, there are also variations depending on how the data is structured. The configurations depend on how the feature space X, the label space Y, and the identifier space I are organized. Figure II.2 depicts a visualization of the possible data distributions.

First, **Horizontal Federated Learning** comprehend the state in which entities that hold the same data structure but the individual identifiers are different. An example of this configuration would be the initial approach of (McMahan et al., 2017) or in (Fernández et al., 2022b RP1, Fernández et al., 2023b RP2, Lee et al., 2023 RP3) where FL is used to train the Google Keyboard recommender system. There all clients share features, but they are, indeed, different clients. Second, **Vertical Federated Learning** represents, for instance, two companies that hold data from the same individuals, but the feature set is different. An example of this configuration would be the browser cookies from different companies, they point to the same user, but they account for differing features set. Lastly, Federated **Transfer Learning** aims to increase data availability through the use of traditional transfer learning between entities that do not share feature or label space.



(a) Horizontal Federated Learning

(b) Vertical Federated Learning

(c) Federated Transfer Learning

Figure II.2: Data partitions in Federated Learning

1.4. Privacy-Preserving Federated Learning

Although the first algorithm introduced by (McMahan et al., 2017) sparked a revolution in how we conceptualize decentralized AI, it has been found to have several drawbacks. One of the main issues is related to clients' data privacy as pointed out by (Zhang et al., 2021). While not sharing the data seems enough to maintain the privacy of users and their behavioral patterns in FL, research has proven that there are still existing attack vectors to FL in general and DL models in particular. For instance, Model Inversion (MI) attacks in which attackers try to recreate the training data used previously (Fredrikson et al., 2015; Geiping et al., 2020) or poisoning attacks in which an attacker pretends to be a legitimate client preventing the model from learning or biasing it to produce inferences that align with the intentions of the attacker (Benmalek et al., 2022).

These attack vectors can be divided into two categories: those related to the nature of the communication network, where an attacker can intercept information transmitted between clients and servers, and attacks that exploit the patterns learned by the model during the training phase.

The issue of transmitting secret information between multiple untrusted parties over an unprotected channel has gained significance in recent years. One such protocol to transmit information without a layer of encryption was initially brought by (Shamir, 1979) where a set of separated and untrusted agents can share a secret (information) by splitting it into multiple shares. In Shamir (1979) the secret is divided in *n* shares and at least *n-1* shares are needed to aggregate and reconstruct the secret and any allocation of less than *n-1* shares cannot reconstruct the secret.

One such secure multiparty computation (SMPC) protocol is SecAgg (Bonawitz et al., 2017) that extended the application of this cryptographic primitive to the context of FL. In SecAgg clients can share their models as if they were shares of a secret with the central server. The central server can then reconstruct the original message upon reception of the shares. The main advantage of this protocol is its minimal impact on model performance, as secret reconstruction does not affect the information as presented in (Fernández et al., 2022b RP1). However, this approach introduces additional communication rounds between clients and the server and among clients for the secret-sharing process.

On the other hand and as mentioned in Bonawitz et al. (2017), despite the advantages SecAgg offers in securing the communication channel, models trained using SecAgg still contain latent patterns that might point to certain aspects in the original training dataset. This issue opens the door for classic MI attacks where an attacker can retrieve substantial information from the training dataset given a trained model (Bagdasaryan et al., 2020; Fredrikson et al., 2015; Shejwalkar and Houmansadr, 2021).

The main way to protect individual contributions, and thus the latent patterns of the in the model is DP. Introduced originally in 2006, Dwork explained DP as a method to ensure the confidentiality of the data when it is retrieved from a dataset (Dwork, 2006). In other words: "*differential privacy addresses the paradox of knowing nothing about an individual while learning useful information about a population*" (Dwork and Roth, 2014). DP as originally defined by Dwork (2006) provides a rather strict guarantee (Wood et al., 2018). The original vision DP offered *binary* privacy, either formally private or not. These strict guarantees limited the usage due to the low trade-off between privacy and utility. To alleviate that, Dwork introduced (ϵ, δ)-DP. ϵ represents privacy budget and determines how much of an individual's privacy a query may lose and by its use, how much it reduces the overall privacy of the system. On those terms, low ϵ represents high levels of privacy (being $\epsilon = 0$ the perfect privacy). On the same page δ represents the probability of information being leaked accidentally.

In formal terms, as defined in Equation II.1, DP is defined as follows: for every pair of inputs X and Y that differ in one row, for every output in S, an adversary should not be able to use the output in S to distinguish between any X and Y. Dwork (2006):*"[DP] ensures that for all adjacent X, Y the absolute value of the privacy lost will be bounded by ϵ with probability at least $1 - \delta$"*

$$\Pr[M(X) \in S] \leq exp(\epsilon) \cdot \Pr[M(Y) \in S] + \delta \tag{II.1}$$

The application of DP involves the addition random noise, typically drawn from a Gaussian or Laplacian distribution (Dwork and Roth, 2014). In FL, the central server adds noise to the combined models before transmitting them back to the clients. This approach allows the specific contributions of the individual clients to be hidden and dissolved within the overall model, thus safeguarding their privacy.

Assuming that all the clients in FL participate in an equal manner in the training, their updates should modify the model from the central server in an equal manner. So, to maintain the different updates within a controlled range, two main clipping strategies have arisen. The name of clipping comes from the fact that the total gradients or sum of gradients are clipped to either a fixed value (fixed clipping) or a changing value (adaptive clipping). The two methods are explained as follows: On the one hand, fixed clipping, which is derived from the original work of (McMahan et al., 2018) clips the values of the updates in a certain predefined range. Although this solution performs satisfactorily, there is a major limitation with regard to the expected clients. Giving the same range to all clients, where variations in data and models are present, seemed an oversimplification of reality. One of the most prominent improvements is adaptive clipping. This strategy is based on the setting of the clipping value in a quantile of the data distribution (Andrew et al., 2021). Adaptive clipping requires some extra noise to combat the privacy lost in precisely calculating the quantiles [2]. Although it will result in more absolute noise, the performance damage that this will cause is hampered by the rapid convergence rate of this method and in general terms it is supposed to perform better. In plain words, the clipping is less restrictive at the beginning of the training where the majority of the "learning" happens to be more restrictive in the final phases where the fine-tunning occurs. We tested such hypothesis in (Fernández et al., 2022b RP1) in which we compared adaptive clipping versus fixed clipping yielding a 9% performance increase for the former.

Adding noise, regardless of the clipping strategy, has a clear detrimental effect on the performance of the model (Fernández et al., 2022b RP1). Higher amounts of noise will result in less performant models and vice-versa. Sometimes, modifying the amount of noise (z) is not enough to strike a positive balance. The result of having a performant model may be at the expense of a low privacy guarantee. To overcome such issues, one option is to reduce the number of clients involved (Q) in every round of training. With few clients participating, it is easier to obfuscate individual contributions, and thus the less absolute noise will be needed. This strategy, while simple, can also lead to low-quality models. A lower value of (Q) challenges the model as it attempts to learn from a smaller pool of available data (clients). This trade-off is reflected in Figure II.3 where high noise and low participation ratio affects almost equally overall performance.

In summary, in the context of privacy-preserving techniques for FL, DP and SecAgg are two widely adopted strategies that form the foundation for securing FL in all its aspects. DP provides a strong guarantee of privacy by adding random noise to the data before sharing them with other parties, thus ensuring that the individual's sensitive information is not disclosed. On the other hand, SecAgg provides a secure way of aggregating the data from multiple parties while ensuring that no individual's data is compromised. These two techniques can be used together. On the one hand, SecAgg secures the aggregated the data from multiple parties. On the other hand, DP adds random noise to the shared secrets, collectively preventing any attacks from both flanks.

[2] Since data is accessed to compute the aforementioned quantiles, some privacy budget is lost in the query, thus some additional noise is required to keep these queries private.

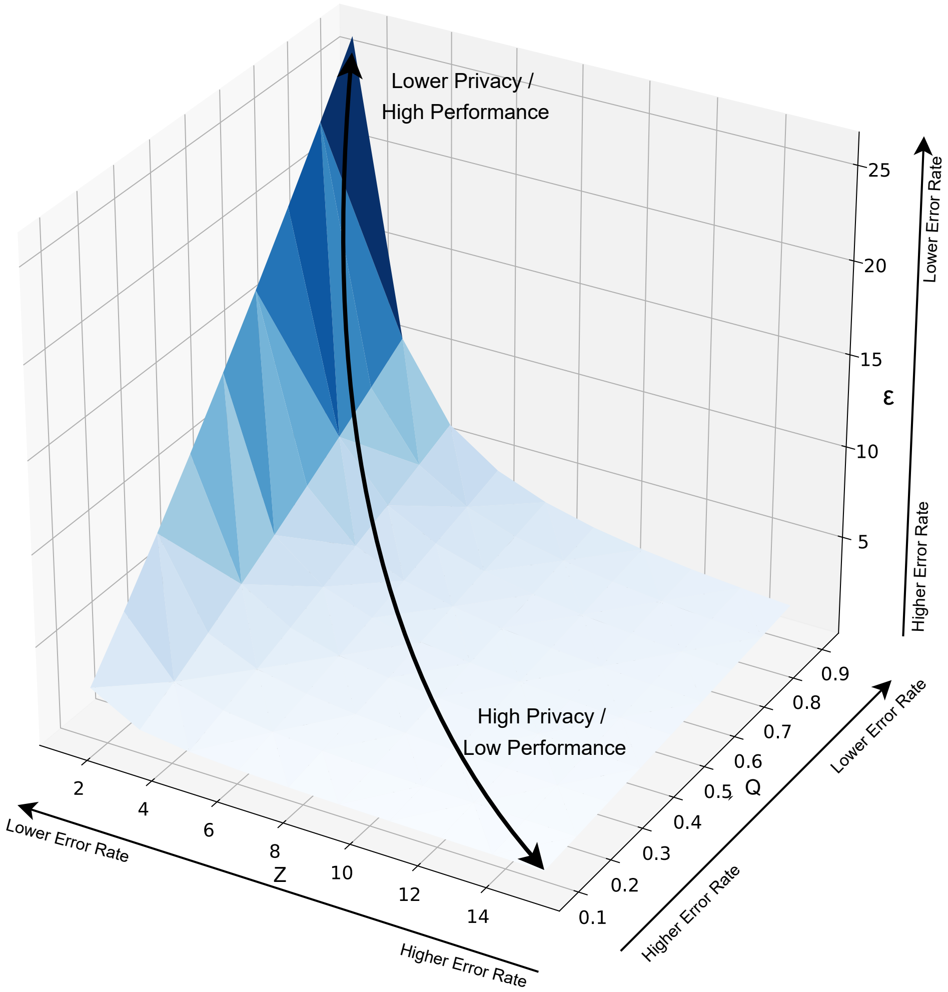


Figure II.3: Illustrative relation between noise (z), ratio of clients per round (Q) and the privacy guarantee (ϵ). The darker the color the more performant the model is.

III | Cooperation amidst competition

1 AI Capabilities: Data is the new gold

The rise in data usage is closely related to the wide adoption of AI, and companies with more data can take advantage of its positive effects, such as gaining a competitive advantage (Mikalef and Gupta, 2021).

Data is often considered a precious commodity and its immense value makes it a highly attractive asset in today's competitive markets (Ciuriak, 2019; Economist, 2017). The benefits of accessing and analyzing data enable companies to make more informed (Merendino et al., 2018) decisions, perform predictive analytics (Shmueli and Koppius, 2011), and personalize their products and services to meet individual consumer service needs (Chen et al., 2012).

Although data is valuable, alone it is not sufficient to provide the insight that businesses require (Mikalef et al., 2018). Data requires new and sophisticated techniques to unlock its full potential (Janiesch et al., 2021). These techniques must be able to efficiently process, analyze and interpret large datasets to reveal meaningful patterns and insights (Manyika et al., 2011). All these techniques fall under the umbrella of AI capabilities as defined by (Mikalef and Gupta, 2021). AI capabilities are the confluence of skills required by a company to choose, coordinate, and use its AI-specific resources to convert data into knowledge, decisions, and actions (Abbasi et al., 2016).

2 The interplay of cooperation and competition

However, sometimes either all AI capabilities or some of them are not always available in sufficient quantity or quality within the company boundaries. One of the main reasons could be data acquisition. Data can be expensive and require advanced capabilities to handle (Aral and Weill, 2007). Another reason may be that access to data is often restricted by organizations for competitive, regulatory, and ethical reasons (Adadi, 2021; Jordan and Mitchell, 2015).

Therefore, given the limitations mentioned above, competitive market participants (e.g., data providers, data aggregators, and public companies) often show reluctance to share proprietary data out of fear of losing an edge (Abbas et al., 2021; Leiponen, 2002). This is clear since the models need data to better generalize the models. Higher model generalization implies a broader view of reality and thus a better ability to understand new and unseen events or data. Thus, companies with better access to data have a broader view of the reality, and a competitive incentive to maintain their data within their boundaries.

For instance, using small and medium-sized enterprises (SMEs) as an example. SMEs are typically characterized by their limited resources and market presence (Bouncken et al., 2015). Hence, they often lack the capacity to accumulate and manage large amounts of data. One potential solution to overcome data limitations is cooperation among SMEs. SMEs can obtain IT resources by collaborating with other businesses. This cooperation helps them to reach market presence, establishing a stronger position. In conclusion, the effect of a cooperation between data-intensive SMEs will alleviate limitations in terms of data ownership and management (Muscio, 2007). In summary, by sharing knowledge and resources (AI capabilities), organizations can achieve common goals more efficiently and effectively.

This stress between cooperation and competition can be tracked from the end of the 20th century era. In 1996, Nalebuff et al. (1996) coined the term "Coopetition"[1] term. It relates to "*simultaneous cooperation and competition between firms*" (Nalebuff et al., 1996). Coopetition posits that firms can improve their performance by collaborating through strategic alliances, networks, and other partnerships. By sharing resources, capabilities, and risks, firms can leverage each other's strengths and mitigate potential weak-

[1] a portmanteau of cooperation and competition

nesses (Bouncken et al., 2015; Hoffmann et al., 2018). This approach can create mutual benefits that lead to greater efficiency, innovation, and competitiveness.

Despite the potential benefits of coopetition, such as increased innovation and market share, companies in competitive industries may view coopetition as a risky endeavor. The challenge for these companies is to find the right balance between competition and cooperation, as "*knowledge leakage, opportunistic behavior, lack of commitment, and instability of interfirm relations*" (Hoffmann et al., 2018) are important concerns to establish coopetitive alliances. Traditional competition scholars have always viewed coopetition as a form of collusion and therefore resulting in an unappealing behavior (Tirole, 1988). In turn, recent research has suggested that such cooperations can bring numerous benefits (Shrader, 2001), which challenges this classical view and supports the idea of coopetition as a viable strategy for companies.

3 Federated Learning as a coopetition enabler

FL is emerging as a promising solution to effectively address the challenges that arise from coopetition. FL acts as a significant catalyst and disruptor of the current coopetitive landscape. As the data is not shared within FL models, it mitigates the risks associated with leakage of business knowledge. This is one of the main concerns in coopetition research (Bouncken et al., 2015). FL can take advantage of isolated training data while keeping it disconnected and private. This paradigm shift is an attractive option for companies seeking to balance competition and cooperation. In this way, FL has the potential to change traditional notions of coopetition, limiting drawbacks such as risking competitive advantage and allowing firms to benefit from cooperation.

The traditional approach to machine learning involves collecting and centralizing large amounts of data in one location, which can create significant privacy concerns for individuals and organizations (Acquisti and Gross, 2006; Dinev et al., 2013). Centralized data storage also introduces security risks as it represents a single point of failure and can compromise the entire dataset. FL, on the other hand, enables entities to collaborate and share knowledge while maintaining data privacy and security.

Although many of the difficulties are not yet fully resolved, FL can improve the way participating entities interact. The coordinating process among clients and the quality assurance of the resulting models are still a subject of interest for the adoption of FL. Regarding the issues related with the nature of coopetition; the risk of knowledge leakage is minimized by design because raw data is not exchanged. Furthermore, similar to the classic collaborative tradition, there are existing incentive methods (Zhan et al., 2020) to prevent opportunistic behavior and motivate companies to engage and ultimately increase commitment. The commitment can be further encouraged by using incentives, such as monetary compensation on improving model accuracy, to motivate participants to contribute their data and computing resources. Finally, because the trust of the model is decentralized and each party retains control over their own data, all players are less dependent on a centralized third party. This, together with the transparency provided by a central shared model, helps promote the trust of the participants (Xu et al., 2022).

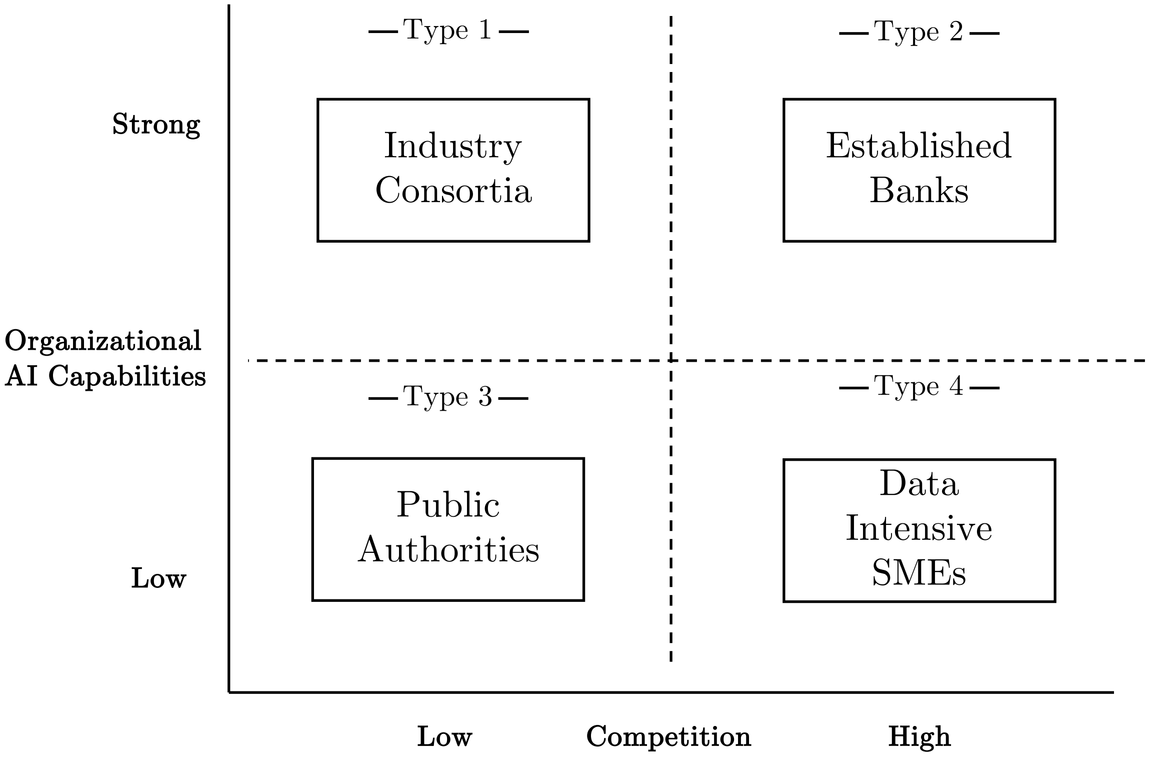


Figure III.1: Conceptualization (with examples) of Federated Learning adoption (c.f. (Fernández et al., 2023a RP 5))

Furthermore, the willingness to collaborate may be related to the nature of the markets (Akhter and Robles, 2006). In (Fernández et al., 2023a RP 5), we postulate that firms with limited AI capabilities, in their attempt to find new prospects, will participate in FL when operating in a competitive market. There, our framework includes two dimensions, on the one hand, AI capabilities within the firm's boundaries; these could be *high* for firms with established market presence and extensive resources, and *low* for small firms with limited access to resources. On the other hand, we model the market dynamics in which these firms participate. This ranges from low competition to high competition. In this framework, we argue that FL has the potential to transform previous competitive dynamics into coopetitive arrangements in which parties can cooperate and reap the benefits of both competition and cooperation. Thus, the purpose of this framework is to map organizations based on their probability or likelihood of adopting FL. There in (Fernández et al., 2023a RP 5) we derived four different types of organizations depending on their market dynamics and the AI capabilities they hold.

Type 1: Strong AI capabilities in low competitive markets The first corner is formed by companies with sufficient in-house capabilities that seek to form consortia in order to achieve a common goal. An example of this is the collaboration between insurance companies, the Alfa association[2]. In this case, French insurance companies join forces to improve their performance in fraud detection. The main distinction between Type 1 and Type 2 organizations lies in the shared objectives that guide the consortium's goals. In Type 1 consortia, cooperation is essential, whereas in Type 2, companies have their own objectives, leading to rivalries and competition. Given the collaborative nature and the shared goals of the consortia, FL appears to be a natural fit for facilitating the process. FL has the ability to handle data heterogeneity and distribution across different organizations. Each organization may have different types of data or data biases, and FL can effectively leverage this diverse data to improve model performance.

Type 2: Strong AI capabilities in high competitive markets The first category, Type 2, encompasses organizations that possess abundant AI capabilities and operate within competitive markets. Examples of such organizations include multinational data brokers, financial institutions, and large energy suppliers (Fernández et al., 2022b RP1, Fernández et al., 2023b RP2).

Due to their strong market position and significant capabilities, these organizations are less inclined to break data silos and share proprietary data. Given their dominance in

[2] https://www.alfa.asso.fr/

the market and access to large databases, they do not find it worthwhile to participate in a FL setting where smaller competitors can benefit from their knowledge and expertise. Instead, they are more inclined to implement FL internally, drawing inspiration from the original proof-of-concept of FL used in the Google Keyboard (GBboard) to enhance predictions across devices (McMahan et al., 2017). This approach allows large companies to leverage their client data while maintaining their privacy.

However, these companies can also utilize their knowledge and resources to further solidify their dominance, potentially pushing smaller players out of the market and leading to the formation of oligopolies. In such scenarios, a small number of large companies dominate the market and utilize their collective power to restrict competition (Osarenkhoe, 2010). This situation can be detrimental to both competition and innovation as it hampers the ability of new and smaller companies to enter the market and challenge established players.

In summary, the adoption of FL by Type 2 companies depends on the specific circumstances. Ethical and benevolent companies are likely to implement FL internally to promote data sharing and improve operational efficiency. On the other hand, malicious actors may exploit the privacy-preserving aspects of FL to collude in the market and gain complete dominance, potentially leading to harmful consequences for competition and innovation.

Type 3: Low AI capabilities in low competitive markets The third corner represents Type 3 organizations whose AI capabilities are limited and which operate in collaborative or low-competitive markets. These organizations typically possess large amounts of data but lack the necessary AI capabilities to effectively manage and train high-performing models with this data. An example of such organizations could be public authorities that lack in-house capabilities to utilize their data and face challenges in externalizing these capabilities due to privacy constraints. Research evidence suggests that a significant proportion of local authorities are hesitant to externalize their processes (Entwistle, 2005).

Furthermore, these organizations may face difficulties in accessing data from other institutions due to privacy concerns, which limits their ability to have a diverse range of data. Collaboration between public bodies is hindered by a lack of optimization and communication, which in turn restricts the sharing of information among network nodes and impedes the flow of information between public bodies (Voskob and Punin, 2003).

In this context, FL offers significant added value. FL allows non-competing entities to leverage their communication flow while minimizing risks, as data remain localized within each organization's silo (Sprenkamp et al., 2023 RP 4). This approach enables organizations to collaborate and learn from each other's data without compromising privacy or data security.

Type 4: Low AI capabilities in high competitive markets The fourth quadrant refers to organizations operating in competitive markets with limited AI capabilities. Examples of such organizations include neo-banks, FinTech startups, and data-intensive SMEs. These entities can overcome their limitations by seeking external sources to acquire the necessary AI capabilities (Winter et al., 2014). FL presents an opportunity for these companies to combine their resources and knowledge, enhancing their AI capabilities and improving their competitiveness in the market. Through coopetition, these organizations will balance their competitive nature with the benefits of collaboration to leverage the collective expertise and data of other participants, enabling them to develop and refine their own AI models without risking their own knowledge.

Participating in FL alliances allows these Type 4 companies to access new AI capabilities or monetize their existing ones. This collaborative effort is particularly advantageous for small sized companies with limited capabilities, as it provides them with the critical mass of data and resources needed to level the playing field and compete more effectively against larger and well-established companies. Therefore, Type 4 companies serve as a prime example of the use of FL.

These companies are characterized by their small size, limited capabilities, and agility to adapt to disruptive digital innovations (Chan et al., 2019). By leveraging FL, they can overcome their limitations and harness the power of collective intelligence to enhance their AI capabilities and remain competitive in the market as presented in (Lee et al., 2023 RP 3) in which small financial institution leverage their knowledge to reach the precision levels of established institutions.

In summary, Type 4 companies operating in competitive markets with limited AI capabilities can take advantage of FL as a means to collaborate, compete, and enhance their AI capabilities. By joining FL alliances, these organizations can access new AI capabilities, leverage their own resources, and effectively compete with larger players in the industry.

Although the risk of data leakage or knowledge is mitigated by using privacy-preserving techniques in conjunction with FL, the mere use of such technology implies the blending and interchanging of AI models. Thus, organizations involved in these partnerships are bound to sacrifice some of their edge.

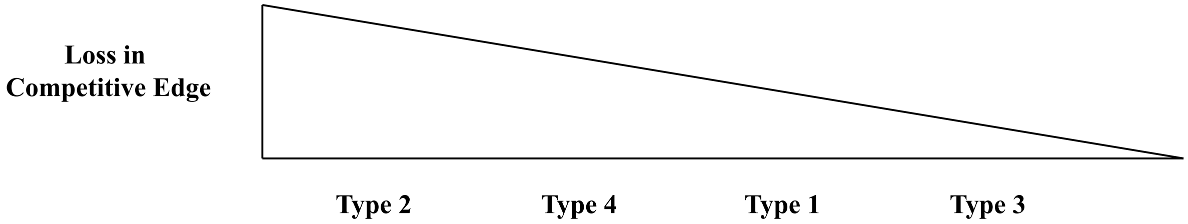


Figure III.2: Graphical representation of loss in the competitive edge.

However, the loss in competitive edge against competitors is not flat (Figure III.2). Obviously, the more competitive the market, the more protected the information is, and thus the more dangerous the disclosure of information. This is evidenced in the work done in (Lee et al., 2023 RP 3) in which Type 2 financial institutions put their expertise at risk by accepting low compensation compared to the increase in the effectiveness of the model. In contrast, small and agile banks (Type 4) take calculated risks to achieve significant gains.

IV | Application of Federated learning

FL can be highly valuable in competitive markets, as it allows entities operating in such markets to join and utilize FL depending on their specific circumstances. However, there are some limitations to using FL, and the following sections examine its suitability in three contrasting scenarios.

Firstly, Section IV.1 investigates the case of large energy suppliers[1] with large AI capabilities under high competitive markets. In such cases, energy suppliers, due to the fact that their data is decentralized, may resort in using FL internally. This is particularly relevant in cases where clients are often physically and logically separated but can benefit from the shared knowledge that FL facilitates. Since their customers are distributed across the energy grid and storing all the data in a central silo might see limitations in terms of data privacy (Eibl et al., 2015), large energy suppliers can leverage consumer patterns to create more precise load forecasts without the need to move and store the data to a central location.

Secondly, Section IV.2 examines the benefits of collaborative modeling in the competitive financial sector and explores the potential of FL to develop accurate decentralized models. The primary objective of this approach is to enable small financial institutions to join forces and achieve market momentum while maintaining the privacy of customer data. Due to stringent regulations such as the General Data Protection Regulation (GDPR), sharing customer data is limited. However, FL can provide a solution to this problem by allowing Type 2 and primarily Type 4 companies to access a more diverse dataset without sharing the actual data. This can lead to improved model performance and a competitive advantage. On the other hand, larger financial institutions (Type 2

[1] Type 2 organizations according to (Fernández et al., 2023a RP 5)

companies) may not derive significant benefits from FL due to their size and existing infrastructure.

Finally, the last Section IV.3 deviates from the focus of the previous two use cases on the private sector by examining the benefits of FL for public administration (Type 3). There, as in the previous cases, privacy assumes a pivotal role in data sharing. Frequently, governments or public institutions face constraints in sharing data and are limited in their capabilities to capitalize on their AI potential. In such situations, the adoption of FL can serve to mitigate issues related to cross-institutional or cross-governmental communication.

1 Federated Learning in the Energy Sector

As the share of renewables increases in the energy mix, so does the price and volatility of energy. This volatility provokes variations in production and consumption, resulting in higher imbalance prices. This is particularly problematic for energy suppliers, as the imbalance prices are multiplied by the imbalance of their entire portfolio, increasing the overall costs. For energy suppliers, errors in their day-ahead forecasts mean more payments to transmission system operators (TSOs) for their imbalances.

Load forecasting and particularly short-term load forecasting (STLF), which aims to forecast the load within a particular short time frame, typically ranging between 1 to 168 hours (Muñoz et al., 2010) is critical for TSOs and energy suppliers. These entities use it to reduce the overall imbalance of their portfolio, thus reducing the imbalance costs. In particular, TSOs are responsible for balancing generation (supply) and load (demand) and must ensure the correct balance between the two. For energy suppliers, a similar operation is required, as they need to meet their day-ahead schedules during their next-day operations. STLF has become especially important at a time when traditional and easily predictable energy sources are scarce and are being replaced by variable renewable energy sources (VRES) that are harder to predict. VRES such as wind or photovoltaic panels introduce higher levels of volatility and imbalances (Muñoz et al., 2010). As a result, high-accuracy load forecasting models have become more important than ever for both energy suppliers and TSOs.

In early 2009, the European Union (EU) implemented the 2009/72/EC (The European Commission, 2009) directive to regulate the installation of smart meter devices throughout Europe. With this regulation, the EU planned to reduce the power system imbalance thanks to a more detailed understanding of household consumption. Furthermore, it offers intelligent and data-driven dynamic pricing for clients that increase the efficiency of the European energy market. The use of smart metering devices has greatly improved the performance of STLF methods by providing more detailed and real-time consumption data. However, the high data granularity obtained with smart meters also raised concerns in terms of accuracy, decentralization and privacy leaked from such detailed energy load profiles (Radovanovic et al., 2022).

The following sections delve into these key concerns. The first Subsection 1.1. focuses on the development of highly accurate forecasting models. Second, Subsection 1.2. presents how FL can leverage the decentralized energy grid. Third, Subsection 1.3. explores centralized and decentralized clustering techniques to address the variability in load profiles and how these clusters can enhance the performance of FL. Lastly, Subsection 1.4. explores and advances the current state-of-the art on privacy preserving techniques for load profiles. As precision and granularity increase, the more revealing and identifiable consumption patterns become, the greater the threat to customer privacy.

1.1. Short-Term Load Forecasting in Power Systems

The challenge of developing accurate forecasting models is exacerbated by the proliferation of smart meters in households. This new availability of data has changed the way forecasting is done. Traditional "*top-down*" involves creating an aggregate profile based on profiles with presumably similar behavior. This approach was fully utilized in times when the cost of imbalance was low. However, as the share of VRES in the energy mix increases, so do the imbalance costs, requiring more accurate energy forecasting. Furthermore, during periods of scarcity of energy resources (such as electricity, oil, and gas), the accuracy of forecasts becomes critical for suppliers. In turn, "*bottom-up*" approaches such as STLF rely on the granularity provided by smart meters and produce individual profile forecasts. The current STLF methods rely on statistical metrics such as moving averages, linear models such as Autoregressive integrated moving average (ARIMA), or in some cases simply comparing load to the previous day.

Recently, to accommodate the extensive data generated by smart meters, new methods have emerged, mostly based on ML such as XGBoost (Bollenbach et al., 2022) or DL, to more accurately predict load curves taking into account the new patterns (Hippert et al., 2001; Nassif et al., 2021; Nti et al., 2020). Although there has been an increase in both usage and research interest in DL techniques, a perfect solution to effectively handle all of the above variations in load profiles has not yet been developed. In particular, research has focused on models that can more accurately identify nonlinear and latent patterns in the data, as opposed to models such as ARIMA, which have limiting assumptions such as linearity or seasonality that hinder their performance.

In (Fernández et al., 2022b RP 1) we reviewed a large body of research on load forecasting models. Our results there suggested that DL models have become deeper likely to capture the nonlinearities in the data. The neural network (NN) layer design found has maintained constant over time with a majority of Fully Connected layers (FCL), Long Short-term Memory (LSTM) Layers (Hochreiter and Schmidhuber, 1997) and convolutional layers. Moreover, looking at the architectures, we found a large utilization of autoencoder architectures. Autoencoders aim to learn a compressed representation of the input data, called a latent space. The bidirectional encoding-decoding phase allows the autoencoder to learn an efficient, data-specific reduction of the input data, where similar inputs will have similar representations in the latent space. This low-dimensional representation helps autoencoders to accurately predict such diverse load profiles. Furthermore, in recent times, more and more "alternative" architectures, such as Attention mechanisms (Sehovac and Grolinger, 2020) or hybrid models (Yan et al., 2018), have been included, deviating from the classic AI tradition of using LSTM for time-series-related problems.

1.2. Federated Learning for Short-Term Load Forecasting

With the increasing penetration of smart metering data in European households, energy suppliers have tried to aggregate them into central forecasting systems. Although forecasting systems have the potential for highly accurate predictions, they face two significant challenges. First, as mentioned above, smart meter data can be linked to individuals, thereby presenting significant privacy concerns. The level of detail captured by smart meters can enable the identification of specific customers. Secondly, there are substantial regulatory disparities that make it difficult to determine data ownership, particularly for smart meter data, since there are no specific regulations in place (European Commission, Directorate-General for Energy et al., (2020); Haney et al., 2009). For instance, under some regulations, data ownership remains with the customer; in other regulations, the energy supplier is the owner. This uncertainty often makes centralized data lake approaches, such as Atrias in Belgium (Atrias, 2021), or Elhub in Norway (Elhub, 2021), less desirable.

FL emerges as a highly suitable solution in scenarios where decentralization is already present, such as the energy grid, where the ownership of smart metering data is distributed across households, offering a promising approach to address the challenges posed by centralized forecasting systems. The underlying principle revolves around treating each household as an independent client within the FL framework, along with a central server, typically represented by the energy supplier. By adopting this approach, individual households can train their own models based on their unique consumption profiles. Subsequently, these households transmitted their model weights to the central server. Performing as a central server, the energy supplier undertakes the crucial task of aggregating and averaging the received models while incorporating privacy-preserving techniques (see Section 1.4.) to protect sensitive information. This integration of privacy measures ensures that the consumption patterns of other households remain concealed from both the energy supplier and other participants. After performing the necessary calculations, the energy supplier returns the updated models to their respective households. Through this iterative process, FL guarantees a collaborative, yet privacy-conscious environment, allowing accurate load forecasting without compromising individual privacy or exposing consumption details to unauthorized entities.

The core of both articles (Fernández et al., 2022b RP1, Fernández et al., 2023b RP2)rely on this argumentation.

Despite its effectiveness, this approach has some limitations that should be taken into account. First, there is a clear trade-off between privacy and accuracy. The empirical findings of our research indicate an average performance decrease of 20% to 40% (Fernández et al., 2022b RP 1). Secondly, inherent variability among households and their load profiles can introduce conflicts during the training phase. If households possess load profiles that push models learning in opposing directions, it can have a detrimental impact on the overall performance of the aggregated model. The diversity of consumption patterns among households can lead to inconsistencies and discrepancies in the training process, which requires careful consideration and mitigation strategies to ensure the effectiveness of the FL framework.

To overcome these challenges, ongoing research aims to develop innovative techniques that strike a better balance between privacy preservation and model accuracy (see Section 1.4.). Additionally, exploring methods to cluster conflicting load profiles within the FL framework holds promise to improve overall performance in decentralized energy grids (see Section 1.3.).

1.3. Clustering load profiles

Clustering households by those with similar characteristics is a technique that can reduce the overall inherent variability of households, thereby leading to improved prediction accuracy (McLoughlin et al., 2015). This reduced variability can ease the learning of the models (Syed et al., 2021). By doing so, similar load profiles (households) can be grouped together, improving the accuracy of forecasting algorithms.

As mentioned above, within FL this problem is particularly important, as the diversity in load profiles can impede the overall learning of the models (Fernández et al., 2022b RP1). Previous studies have confirmed the effectiveness of centralized clustering before training FL models to cope with the variability. Nonetheless, as pointed out by (Saxena et al., 2017), load profile clustering typically necessitates global data access, which entails a centralized framework. This global access contradicts the decentralized nature of FL. As a result, the clustering techniques and FL encounters incompatibilities, as in (Han et al., 2020; He et al., 2021; Savi and Olivadese, 2021) and (Fernández et al., 2022b RP1).

Ideally, these clusters will be generated *ad-hoc* by the clients as time progresses to adapt to the changing load profiles of customers (e.g. someone buying an Electric vehicle).

In (Fernández et al., 2023b RP 2) we proposed a decentralized P2P clustering approach for households utilizing Agent Based Modeling (ABM) principles. Our goal is to generate clusters in a decentralized P2P manner, eliminating the need for a central server. In this approach, agents, represented by their load profiles, have the freedom to select the most similar agents and form clusters. By creating these clusters, we can reduce variance across households, improving thereby the performance of load forecasting.

Later, these groups will eventually form federations in which FL can train with reduced variability, thus increasing the performance of the model. Calculating the similarity or distance between load profiles is a fundamental requirement for any clustering technique. Classic clustering techniques, such as K-means, use Euclidean distances as their way to calculate the similarity. However, clustering load profiles is not a trivial task(Fernández et al., 2022b RP1, Fernández et al., 2023b RP2). Issues with alignment can occur on multiple occasions, modifying the overall similarity between two load profiles and consequently impacting the assigned clusters. Traditional metrics such as Euclidean distances struggle to solve these alignment issues.

Euclidean distances compare time-series point-to-point, ignoring misalignments or shifts. This means that if two time-series are in-phase, the Euclidean measurement fails to capture the distinctive characteristics of each individual time-series (Fernández et al., 2023b RP2). In such cases, more advanced metrics like Dynamic time warping (DTW) prove to be effective in mitigating alignment issues. DTW aligns the sequences to identify the optimal match between them, even if they exhibit variations in length, speed, or non-linear behavior. Although DTW is a powerful technique, it is also computationally expensive, on the order of computational complexity $O(n^2)$.

$$dtw(i,j) = \begin{cases} \infty & \text{if } i = 0 \text{ or } j = 0 \\ 0 & \text{if } i = j = 0 \\ \|X_i, Y_j\|_2 + \min \begin{cases} dtw(i-1,j) \\ dtw(i,j-1) \\ dtw(i-1,j-1) \end{cases} & \begin{matrix} i = 1,...,n & X \in \mathbb{R}^n \\ j = 1,...,m & Y \in \mathbb{R}^m \end{matrix} \end{cases} \tag{IV.1}$$

$$d(X,Y) = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \tag{IV.2}$$

Equations IV.1 and IV.2[2] represent the disparity in complexity mentioned above. On the other hand, Figure IV.1, displays the visual differences between the DTW and Euclidean distances for the same two household load profiles over 48 hours. DTW finds alignments across the spikes between $t = 15$ and $t = 20$, however, Euclidean does not, as demonstrated at $t = 42$, where the spike in the above profile is measured against a valley in the below profile.

However, it remains unclear whether the use of advanced distance metrics such as DTW, despite being computationally intensive, yields superior results compared to simpler and easier-to-calculate metrics such as Euclidean distances. Our experiments revealed that DTW required significantly more time for the calculations without sub-

[2] Assuming that $X, Y \in \mathbb{R}^n$

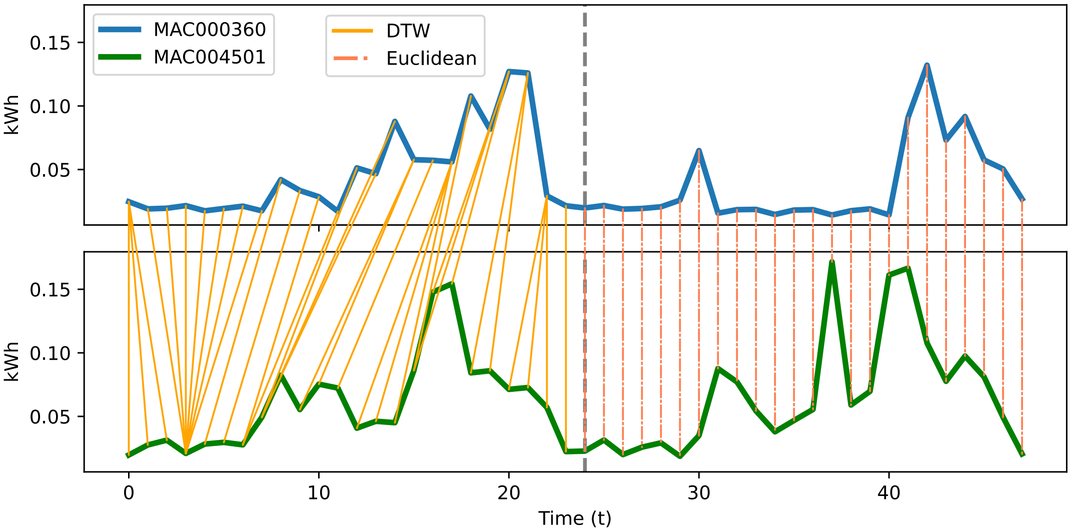


Figure IV.1: Comparison of Dynamic time warping (left) and Euclidean (right) distance over 48h for two residential load profiles. (c.f. (Fernández et al., 2023b RP 2))

stantial performance improvements (Fernández et al., 2023b RP2). The marginal impact of performance associated with this approach can be attributed to the inherent characteristics of the models employed. Once outliers are effectively isolated, these models gain a sufficient understanding of consumption patterns, reducing the need for highly accurate clusters provided by DTW. In (Fernández et al., 2022b RP 1) we could already foresee a similar behavior in which, by simpler Pearson correlations between households, we could improve the overall forecast performance more than 15% on average.

Based on our findings, we reached the conclusion that when combined with a P2P clustering approach, Euclidean distances can effectively generate clusters suitable for FL, enabling accurate and efficient learning of load patterns without the need for complex alignment operations required by DTW. Furthermore, this approach eliminates the requirement of aggregating data in a central repository, fostering a more decentralized and privacy-preserving framework.

1.4. Privacy-preserving federated learning for short-term load forecasting

The more detailed the available data, the more likely specific characteristics of the households will emerge, making them identifiable (Radovanovic et al., 2022). Privacy concerns have arisen with the availability of high-resolution smart meter data, as it provides insight into customer behavior (Fan et al., 2013; Kim et al., 2011; Kolter and Jaakkola, 2012; Wang, 2020). Studies have successfully used load profiles to detect holidays (Eibl et al., 2019) using 15-minute resolution load profiles from Upper Austria. Although determining the occupancy status of a household may seem straightforward, research has progressed to identify specific characteristics of households, such as detecting swimming pools (Burkhart et al., 2018; Ferner et al., 2019) or the presence of air conditioners (Pathak et al., 2018). These efforts highlight ongoing efforts to uncover unique attributes and behaviors within load profiles for a deeper understanding of customer usage patterns.

In FL, although raw data is not shared, models are continuously exchanged. Therefore, privacy has become a crucial issue, particularly when forecasting future load consumption can reveal individual characteristics. DP has emerged as a prominent method in FL to ensure privacy preservation, as mentioned in Section 1.4.. However, determining the appropriate level of noise to achieve privacy protection remains an open research question (Hsu et al., 2014; Kohli and Laskowski, 2018; Lee and Clifton, 2011), and so does in the context of STLF.

Simply adding noise to the data does not guarantee that privacy requirements will be met in all scenarios and vice versa. Privacy requirements in domains such as health data are fundamentally different from other privacy standards, and require much stronger privacy guarantees to avoid catastrophic outcomes (Kaissis et al., 2021). Although the privacy requirements for load data are not as high (Eibl et al., 2018), they should not be ignored, as they contain personal information that must be protected.

Due to the fragility and sensitivity of STLF models to noise perturbations, there is a fine line between privacy and performance, making it essential for STLF models to determine the right level of noise (Fernández et al., 2022b RP 1). Normally, more privacy means more noise, which is bad for performance and vice versa. The performance will increase with less noise, but there will be less privacy.

In (Fernández et al., 2022b RP 1) we aim to find the optimal DP noise level for STLF within a FL application. There, we try to find the balance between maximizing the noise level and minimizing its impact on the performance of the forecasting model. Thus, we performed a *grid-search* on the privacy hyperparameter spaces to train different models under a large range of different differential privacy requirements (see Figure II.3). There, the aim is to obtain the most private model possible given the performance benchmark. In turn, this approach may add more noise than necessary to ensure the privacy of individuals. It is thus likely that a less private model may still be sufficient to defend the model against attackers.

From the work performed in (Fernández et al., 2022b RP 1) we can extract some design principles to successfully train federated STLF models using DP. First, the retraining of private models. The idea of this retraining lies in its simplicity. Models trained with DP can be difficult to fit under such compromising constraints; sometimes the optimizations mentioned above produce either not enough privacy, not enough performing models, or there are not enough clients to converge. One way to deal with this is to retrain the DP model for each of the respective clients. Since the DP models are private, retraining for a specific client will still maintain privacy with respect to itself. The results of the analysis show a significant improvement in the accuracy of the prediction with minimal performance compromise.

Second, it also proposes some guidelines for the model architecture to properly deal with noise. The sensitivity of predictive models to noise depends not only on the amount of noise or its optimization, but also on their internal architecture. In general, STLF neural architectures tend to rely heavily on autoencoders (Khan et al., 2020; Marino et al., 2016; Sehovac and Grolinger, 2020) and (Fernández et al., 2022b RP 1). Autoencoders are well suited for FL because they can adapt to the variability in data between different clients. However, our research has shown that these models are more sensitive to noise than traditional stacked LSTM networks. This increased sensitivity may be due to the fact that autoencoders compress information from high-dimensional spaces into a lower-dimensional representation. Following the work of (McMahan et al., 2018), noise is added to the model weights after aggregation. There, we found evidence that the noise added to the latent space perturbs the *mapping* between inputs and outputs. This perturbation amplifies the effect of the noise because the model cannot find the connection between the encoded and decoded features during the decoding phase, preventing the model from learning any of the patterns.

Lastly, as mentioned in (Fernández et al., 2022b RP 1) the use of SecAgg appears to cover the gaps left by DP. Although models can be *attacked* using MI attacks (Bagdasaryan et al., 2020; Fredrikson et al., 2015; Shejwalkar and Houmansadr, 2021), SecAgg as it is built on SMPC protocols offers strong privacy guarantees to share model weights or gradients efficiently and securely.

Although the original work pursued in (Fernández et al., 2022b RP 1) aimed to find optimal levels of noise to ensure high performance private forecasting models. We did not study whether such privacy constraints will protect the households from attacks or whether lower levels of privacy are enough to protect individuals without hampering the training.

Herewith, the articles (Fernández et al., 2022b RP1, Fernández et al., 2023b RP2) aim to close ties with the utility of this data in a decentralized manner, offering methods and a self-sufficient technological stack to produce accurate yet private forecasts. On the one hand, providing methods to mitigate attack vectors on load profiles while maintaining the trade-off between privacy and accuracy. On the other hand, to propose new P2P mechanisms that allow households to form groups with individuals who share similar characteristics to mitigate the significant variability introduced by smart metering devices.

2 Federated Learning in the Financial Sector

Financial institutions regularly assess the creditworthiness of individuals and entities seeking loans or credit. This process is crucial to determining an institution's exposure to risk. Through the gathering and analysis of both quantitative and qualitative data, financial institutions are able to make informed predictions about a borrower's ability to fulfill their financial obligations in the future. Traditional methods rely on principles of induction to make mathematical and statistical inferences from curated data. These inferences are limited by assumptions such as linearity, independence, and normality. To overcome those limitations, modern intelligent approaches use computational methods that rely on patterns extracted from data and thus decouple them from the mentioned assumptions (Chen et al., 2016; Galindo and Tamayo, 2000). Therefore, credit risk models created from intelligent methods perform better in many cases by generalizing complex real-world data where noise, non-linearity, and idiosyncrasies are observed regularly.

Previous empirical studies have shown that modern credit risk models sometimes perform better than those created by traditional methods, but even such credit risk models are constrained by the same limitations. The performance of credit risk models depends not only on the methodology used but also on the data input (Altman, 2002; Heitfield, 2009).

Access to data is crucial for the effectiveness of intelligent methods, as they require large amounts of data to be accurate and reliable in uncertain situations. In particular, in the mortgage market, where data on credit risk and, by extension, mortgage risk, are costly, scarce (Jha et al., 2012) and subject to stringent regulations such as, for example, General Data Protection Regulation (GDPR) in Europe, which limits the general usability of the data. Consequently, having data generally offers a competitive advantage to its holder (Lee et al., 2023 RP 3) (Bansal et al., 1993; Walczak, 2001). Furthermore, the quantity and quality of these data can significantly affect the performance of credit risk models (Bansal et al., 1993; Walczak, 2001).

The efforts to increase data availability and increase performance of these models are impeded by data privacy issues and restrictions on data sharing and collaboration (Borgman, 2012; Ekbia et al., 2015). Concerns about data privacy are pronounced when the sharing of data in question is sensitive, valuable, and across silos as data allows for precise identification of individuals. On the other hand, data also require investment to produce and maintain as a corporation. The competitive landscape pressures corporations to be reluctant to openly share data.

Data availability and privacy are two significant challenges that can limit the effectiveness of credit risk models. To overcome these limitations, an approach is to use FL to form cross-organizational agreements that provide access to new AI capabilities or enable the monetization of existing capabilities (Fernández et al., 2023a RP 5).

"*Financial institutions would not need to reveal their data as they gain insights from its processing, allowing every participating Financial Institution to benefit from the use of each other's information*" (Lee et al., 2023 RP 3). By employing FL, financial institutions can create joint credit risk models without sharing raw data. For example, small financial institutions, such as FinTech startups and neobanks that need to purchase data from external sources to stay competitive, can leverage the knowledge of others, broadening the scope of their models. As a result, these small financial institutions can lower its internal bias and better accommodate new customers, improving their overall performance on credit-related tasks.

This is particularly important for smaller financial institutions, such as FinTech startups or neobanks, who need to acquire data from external sources to remain competitive. In (Lee et al., 2023 RP 3) we explored different scenarios to analyze the performance and utility of FL and credit risk assessment. To evaluate the suitability of this model for financial institutions, we collected data on American mortgages over a three-year period. The original dataset comes from Freddie Mac's Single Family Loan-Level dataset (SFLD) (Freddie Mac, 2021) that contains up to 9M mortgage observations. An observation is defined by the state in which a mortgage is. Given a month, we have access to the internal state of all active loans under their management.

What makes this data set particularly interesting is that not all banks had the same amount of mortgage under management, allowing us to test both the hypothesis of knowledge sharing between small financial institutions to acquire AI capacities to better compete (Bouncken et al., 2015) and the importance of large financial institutions relative to the small ones in a model sharing relationship.

To evaluate our initial assumptions, we built a list of scenarios. Initially, we evaluated the performance of financial institutions just by estimating their risk based on their own data; this scenario serves as a realistic baseline for future comparisons. Second, we have a hypothetical scenario formed by a large data lake with all the mortgages. Third, there is an FL scenario where all banks participate equally in this form of coopetition. Fourth, and to evaluate the hypothesis where we posit that small financial institutions have an incentive join FL, we model two more scenarios where the most prominent bank in America and the second most prominent banks are excluded from FL. With these scenarios, we sought to evaluate the importance of the largest mortgage transactor with respect to the knowledge that a smaller version of these alliances might have.

Our simulations suggested that first, when all financial institutions join a data lake, the evaluation metrics were close to 100%. This initial result seems intuitive but impractical. There, we assume that all financial institutions within a jurisdiction will openly share their data with a central entity for future training. Second, we tested the main hypothesis of the article, *is it enough for small financial institutions to just collaborate between themselves?* As previously stated, we validated this hypothesis by training two distinct and independent FL models, one of which excludes the top financial institution while the second one, building on the previous, also excludes the second largest financial institution. Our findings reveal that the results of both models are in pair with the performance obtained in the scenario where all financial institutions participate and even in the initial data lake scenario where all the data are aggregated in a silo.

Furthermore, our simulations revealed that by collaboration financial institutions can not only better estimate the risk of default of a mortgage, but also anticipate unforeseen events of the mortgage. In Figure IV.2, we explore the probability of default estimated by any of the banks and FL for a particular loan. In this case, we can visualize how the smallest financial institution in the dataset (Metlife Bank) reacts late, while at the end of 2012 the rest of the financial institutions (and FL) gave this loan a high default rate, Metlife Bank needed at least one extra year to realize that the likelihood of this mortgage to be paid-off was minimal. As seen, our experiments suggest that small financial institutions can profit from FL in two ways: Firstly, better anticipating unseen events and secondly, having a higher accuracy when estimating the default likelihood of a mortgage.

This behavior can be attributed to the fact that the credit risk data are tailored to both the issuer and the financial institutions, and thus have a limited scope for estimating the overall risk. The data they have is not sufficient for a robust assessment of the overall risk distribution (Taleb, 2020).

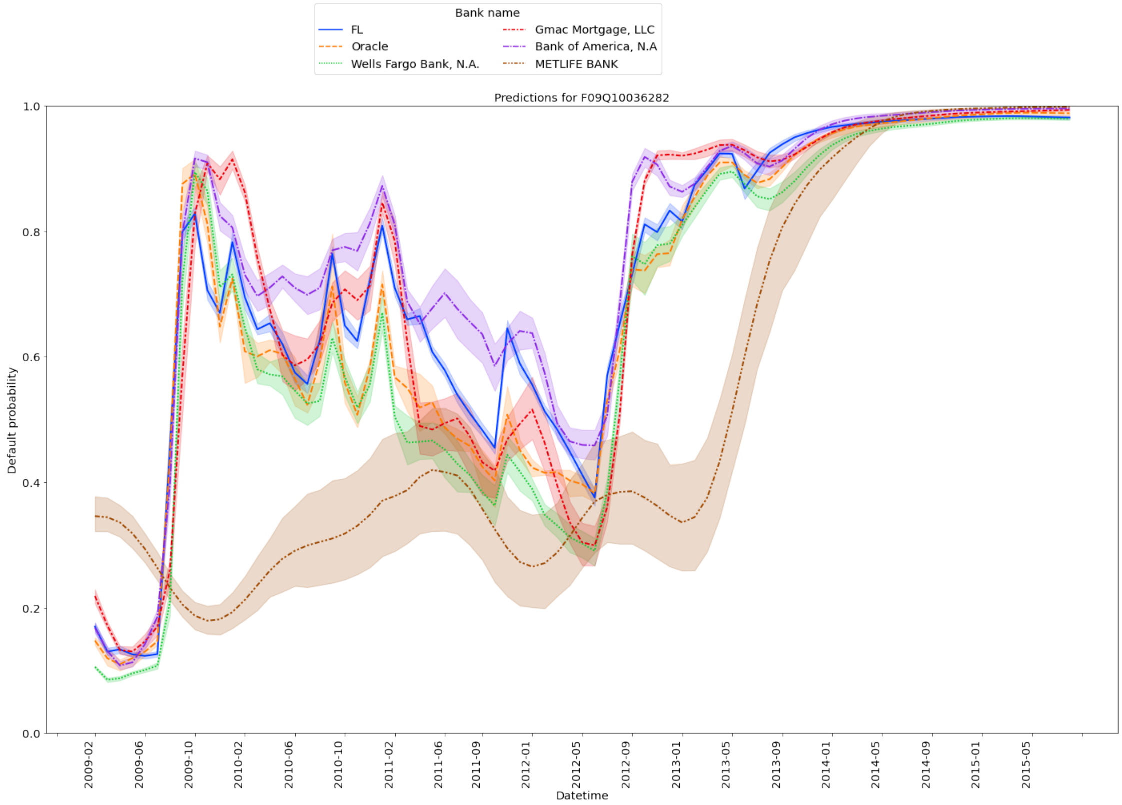


Figure IV.2: Reaction capacity between financial institutions

In summary, the case of financial institutions in the context of credit risk assessment is clear, as a relatively small financial institution could benefit significantly from collaboration with others through FL. This particular case might be because every financial institution holds data for a specific market niche. Collaboration allows smaller companies to combine their resources and expertise to access new markets and gain access to knowledge that would otherwise be reserved for larger institutions. This can lead to increased competitiveness and growth for smaller companies.

In addition to the empirical benefits that this technology can bring to the financial sector, there are also counterparts to take into account. One of them could be the formation of alliances or oligopolies. These oligopolies are usually formed by competitive entities in which they find significant benefits by collaborating rather than competing. Securing their marked dominance by doing so (Goyal and Joshi, 2003). For example, consider an alliance of top financial institutions that collaborate to develop FL models for more accurate loan calculations. In this scenario, smaller financial institutions are considered insignificant and excluded, whereas larger financial institutions end up dominating the market. This type of environment is conducive to the formation of data cartels and oligopolies, as only large organizations can form strategic alliances with each other based on their significant contributions (Goyal and Joshi, 2003).

3 Federated Learning in the Public Sector

Governments are currently not prepared to harness the potential of the rapidly growing field of AI (Wirtz et al., 2020). This new technology-driven wave has been primarily driven by the private sector. However, practices that are effective in the private sector may not be applicable in the public sector due to differences in priorities. While the private sector often prioritizes competitive advantage, the public sector is more concerned with maximizing public value (Fatima et al., 2020; Zuiderwijk et al., 2021). Consequently, there is less AI knowledge and expertise within the public sector compared to the private sector.

This new technological wave describes the increasing use of disruptive information communication technologies (ICTs) such as AI in governments. The so-called e-Government 3.0 aims to use AI techniques to solve societal problems (Sprenkamp et al., 2023 RP 4).

However, obtaining the necessary data for this initiative is not easy. The government technological infrastructure is limited and there is a shortage of AI expertise in-house among public administrations (Ojo and Millard, 2017). Additionally, government data may have explicit privacy concerns (Isaak and Hanna, 2018), which makes data acquisition even more challenging.

An option for governments is to acquire their IT and AI capabilities through intergovernmental data sharing (Sprenkamp et al., 2023 RP 4). This approach is similar to how companies seek the necessary capabilities beyond their organizational boundaries (Winter et al., 2014), where companies without the necessary IT capabilities to run projects *in-house* look for services outside the limits of their own company.

Addressing the data hunger of AI models and resolving the obstacles faced by intergovernmental alliances can be achieved through the sharing of data by (Thiebes et al., 2021). However, several challenges arise in the context of such alliances. These include resistance from organizations to data sharing (Gupta, 2019; Sun and Medaglia, 2019), handling data with special privacy rights, such as personal information (Isaak and Hanna, 2018) or the consequences of sharing data in a field where there is no shared legal framework.

In (Sprenkamp et al., 2023 RP 4) we focused on the practical issues of intergovernmental data sharing and the potential role of FL in addressing them. FL offers a promising solution for governments that require large amounts of data, but face constraints in terms of financial incentives. Allows for decentralized model training without the need to aggregate or share data centrally. This approach has the potential to address privacy concerns and facilitate data sharing between governments, especially in contexts where data protection laws are stringent or negotiating data sharing agreements is challenging.

Explicitly, these practical issues are mentioned by the Organization for Economic Cooperation and Development (OECD) in OECD (2019). First, those related to the conflict of interest between the parties involved in a data sharing agreement, to this end, FL can offer security and privacy towards raw data, as well as inherent protection against violation of intellectual property. Second, thus with regard to the trust and re-use of data across societies. These mentioned topics involve supporting communities and building standards with a particular focus on data quality. These issues are solved by FL since although the technology is recent, there are already methods to deal with data quality (Passerat-Palmbach et al., 2020) or system heterogeneity (Zhang et al., 2021) . The last problem mentioned by the OECD relates to misaligned incentives and limited data sharing business models. In this case, FL is prone to opening new business possibilities. For example, the work of (Kang et al., 2019; Yu et al., 2020) delineates the possible incentive mechanisms of FL. On the same page (Balta et al., 2021) propose the business model “*which gives entities an incentive to take part in intergovernmental projects.*” (Olson, 1965) This helps to overcome the disincentives that arise from "free riding" and collective action problems, which can lead to inefficiencies in data sharing (Sprenkamp et al., 2023 RP 4).

4 Positioning the applications

In (Fernández et al., 2023a RP 5), we presented a framework that allows us to analyze how companies position themselves based on their vision of FL. Similarly, as presented in Figure IV.3, the papers included in this dissertation can also be classified within this framework.

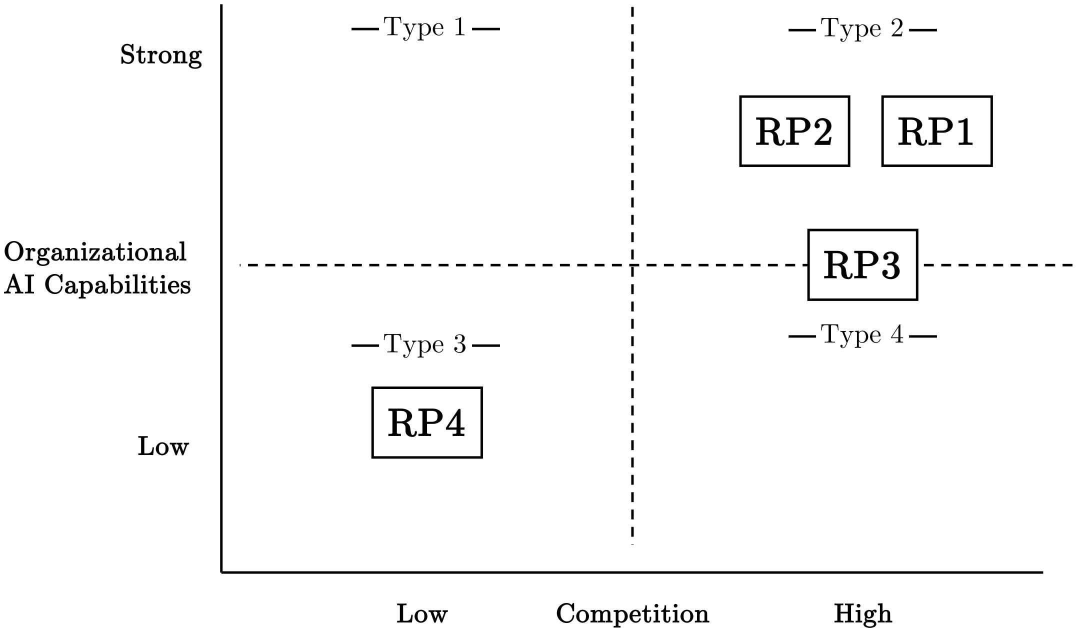


Figure IV.3: Placement of all research papers within the (Fernández et al., 2023a RP 5) framework.

The proposed framework can initially be divided into four quadrants, in which examples of Type 2, Type 3 and partially Type 4 companies are predominant in this dissertation. The first quadrant that collects the majority of the articles is the top right, Type 2 companies describe situations where companies that typically have enough capabilities within their boundaries, coupled with a highly competitive market. These companies will seek to deploy FL in-house. The work conducted in energy sector around household load profiles done in (Fernández et al., 2022b RP1, Fernández et al., 2023b RP2) falls within the quadrant.

The use cases presented in both papers involve energy suppliers with abundant AI capabilities but who do not own the data of their customers. The customers themselves are the sole owners of their data consumption. This, together with the dispersion across the energy grid, difficulties the creation of centralized models. Energy suppliers can utilize FL to take advantage of the distributed data ownership to create more precise load forecasts without the need to move, store or purchase the data.

The second predominant quadrant is the one related to entities with low capabilities and low competition regimes. Type 3 companies, such as public bodies, can use FL to streamline processes and share technological advances. The work done in (Sprenkamp et al., 2023 RP 4) refers to such a scenario. There, we evaluated the potential fit of FL within the public sector. By using it, governments can take advantage of technology to streamline their processes. The governance literature has previously identified limitations in the IT and AI capabilities of public institutions, due to the limited knowledge of machine learning among personnel and the high financial cost of complex infrastructure projects (Ojo and Millard, 2017). For example, building a sophisticated technological infrastructure to store and collect data would require significant investment (Wirtz et al., 2020). However, the use of FL can help overcome these challenges by reducing the need for complex infrastructure, as the amount of data required to store data and models is minimized with the use of FL technology.

Lastly, the work performed in (Lee et al., 2023 RP 3) can be categorized into Type 2 and Type 4 companies. This article explores the potential collaboration among financial institutions that have traditionally been competitors in the mortgage market, with the aim of enhancing their credit risk models and overall performance. It is important to note that large financial institutions (Type 2) possessing substantial capabilities are impacted differently compared to small fintech startups or neo-banks (Type 4). The former may develop oligopolies through cooperation, while the latter may gain market presence through coopetition.

As mentioned above, the paper explores both scenarios. Type 2 banks, which have an extensive network of loans under management, do not derive significant benefits from FL. On the other hand, Type 4 banks, although smaller in size, have the opportunity to join forces, leverage their knowledge and data, and effectively compete with larger entities. This highlights the potential for smaller banks to utilize cooperative strategies and collaborative models to enhance their competitiveness in the market.

V | Conclusions

1 Challenges and limitations

Although "*All that glisters is not gold*" (Shakespeare, 1598), the benefits of FL for companies with stringent data requirements are obvious. With the benefits come the limitations. First and foremost, as with any AI model, FL models are data hungry; FL models require large amounts of data to train, and while standard centralized ML models collect all of this data from a single location, the inherent decentralization brings with it some challenges related to the various data sources from which the data is derived. The presence of different data sources inherently creates heterogeneity, which can be characterized not only by differences in data structure, but also by differences in cleaning, labeling, and the inherent stability of the client's network.

In addition to the technical challenges of FL, there are also limitations to the adoption of FL. As mentioned throughout this dissertation, FL appears to be a catalyst for coopetition, where competing firms collaborate to increase their access to data and thus increase their efficiency in data-driven decision making. Although the benefits for small firms in the search for data are obvious, the outcome for large firms is different, as there is a significant risk of collusion in the interplay between cooperation and competition, forming cartels and oligopolies and thus excluding smaller competitors.

The success of this technology is closely related with its adoption and is intrinsically linked to the cost-benefit of participating in such efforts. In other words, for FL to be adopted, participants' contributions must be fairly compensated through incentive mechanisms.

The accountability of the models to ensure contribution share becomes even more critical when all participants aim to optimize the model for their individual benefit in FL. In collaborative environments, two key problems related to incentives stand out. Firstly, self-interested participants like free-riders and biased individuals can jeopardize collaboration's success. They benefit from the shared model without making valuable contributions, either by not adding novel data, fabricating data, or providing false information.

The second problem pertains to the fact that the long-term success of the FL environment is contingent on the continued participation of the data owners in the model training (Bi et al., 2023). As their gradual contributions are used to build a shared model which in turn generates utility, there is a temporal mismatch between training of models (contribution) and commercialization (rewards).

In addition, FL has the potential to obscure the workings of a conventional AI system. While privacy-preserving techniques and the federated nature are advantageous from a coopetitive market perspective, increased privatization of inputs have detrimental effects on auditability, and thus on the accountability of the model challenging fair compensations.

2 Contributions and outlook

FL proves to be an effective technology for overcoming data constraints. FL offers new alternatives to share knowledge between companies without the need to share raw data between them. This feature opens the door to unlimited coopetitive relationships between companies from which they all benefit.

RP 1 aims to thoughtfully explore the positive aspects that FL can bring to energy suppliers in their current day-to-day business. The flood of smart metering devices in households has brought unlimited amounts of data. Although beneficial, there are still nudges to tackle such as privacy of individuals or hard-to-forecast profiles.

From this paper we can expand several design principles for FL in the context of Short-Term load forecasting. For instance, due to the high risk of overfitting, it first recommends a *shallow* prediction architecture. Second, it advises against using autoencoders because DP-FedAvg breaks the connection between inputs and outputs. Third, it encourages the use of Adaptive-DP for both usability and performance reasons. Finally, because of its low performance impact, it emphasizes the use of SecAgg to secure the communication channel between clients and server. In general terms, FL seems to be successful in electric power systems enabling *"high level of information sharing while ensuring privacy of both the processed load data and forecasting models"* (Fernández et al., 2022b RP 1).

RP 2 extends RP 1. In RP 1, among other questions, we introduce problems in learning short-term prediction models in highly volatile scenarios. These scenarios arise because all smart meter data is treated the same, and thus different types of customers (electric vehicle users or people in rural areas) are not taken into account. In this paper, we investigated clustering algorithms for FL. First, we analyzed the need for complex distance metrics, such as DTW, to create adequate time series clusters for FL, and second, the suitability of P2P clustering algorithms. Our results suggest several design principles in this context. First, there is a performance gain when clustering is done before training. Second, it is advisable not to use DTW, since simpler Euclidean metrics outperform it both in terms of computation time and performance. Third, peer-to-peer algorithms seem to be a suitable solution for short-term FL prediction in cases where decentralization is needed, since both the performance and computational overhead trade-offs are negligible. More pointedly: *"Our results reveal the possibility of using P2P clustering along with simple Euclidean distances and FL to obtain highly performant load forecasting models in a fully decentralized setting."* (Fernández et al., 2023b RP 2).

RP 3 examines the performance of credit risk assessment based on FL. In this paper, financial institutions have the ability to estimate credit risk using information not only from individually sourced data, but also from information derived from data from other financial institutions without explicit access to the data itself. Two conclusions can be drawn from this paper. On the one hand, the fruitful cooperation among financial institutions in building credit risk models can increase the accuracy of the models almost to perfection. secondly, the models still perform well even if the largest mortgage holders are excluded from the training and only small financial institutions participate. In other words: *"smaller financial institutions can expect a significant performance increase in their credit risk assessment models by using collaborative machine learning"* (Lee et al., 2023 RP 3).

RP 4 explores the difficulties associated with data sharing between governments and proposes the use of FL as a potential solution to these challenges. Our findings suggest that FL could partially or fully overcome many of the problems outlined by the OECD, making it a promising avenue for future research in the public sector. Furthermore, in cases where data sharing regulations are "*fragmented regulatory landscape*" (Sprenkamp et al., 2023 RP 4), FL appears to alleviate data sharing tensions.

RP 5 *de-facto* summarizes the previous work of this dissertation. In this paper, we focus primarily on the transfer of knowledge from computer science to information systems and the limitations to the adoption of this technology, in particular on "*how, beyond the technical realm, there are barriers to adoption.*" (Fernández et al., 2023a RP 5). To achieve this, we reviewed the current literature to draw conclusions and limitations for the application of FL in an intra- and cross-organizational context. By studying the implications of FL in two different contexts, namely organizational and environmental, we draw a conceptualization model to understand the situation that entities will be entitled to when adopting FL. In addition, we propose a concise research agenda at the intersection of FL research and information systems to promote the adoption of the technology.

In summary, this dissertation extends and investigates the challenges and potential solutions in various domains. It explores the application of FL in energy systems, emphasizing its ability to facilitate information sharing while ensuring privacy. This dissertation proposes design principles for FL, including the use of shallow prediction architectures and privacy-preserving techniques such as differential privacy and secure aggregation attempting to strike a balance between accurate load forecasting and preserving individual privacy rights. These methods minimize the risk of reidentification or unauthorized access by ensuring that sensitive data used in the forecasting process remain anonymous and secure. In addition, the thesis explores the benefits of FL clustering algorithms in volatile electric load forecasting scenarios. It suggests that peer-to-peer algorithms are suitable for FL environments, providing high performance with minimal computational overhead.

Furthermore, this dissertation provides valuable insights into the applications and design principles of FL in different domains, namely the financial and public sectors. It highlights the potential of FL to overcome data limitations, ensure privacy, and improve performance by demonstrating the benefits of collaboration and promoting data sharing.

This research contributes to the advancement of FL technology and provides a foundation for further exploration and its adoption in relevant industries.

3 Recognition of previous and related work

This dissertation is genuinely inspired by the "*extraordinary atmosphere of creativity, intellectual effervescence, openness and friendship he fostered around him*" (Houellebecq and Wynne, 2001, on Niels Bohr's research institute). For me, this occurred as part of the Digital Financial Services and Cross-Organisational Digital Transformations (FINATRAX) research group of the University of Luxembourg's Interdisciplinary Centre for Security, Reliability and Trust (SnT).

I coauthor RP 1,RP 2 ,RP 3 ,RP 5 with colleagues from FINATRAX. Lastly, I coauthored RP 4 together with the Digital Society Initiative within the Information Management research group from the University of Zurich.

Prominently, the articles in this dissertation rely on the solid decentralization foundations of FINATRAX (Albrecht et al., 2018; Barbereau et al., 2023; Rückel et al., 2022; Schlatt et al., 2023; Sedlmeir et al., 2020). This decentralization stream has inspired and guided all the articles thereafter presented. The importance of energy research within FINATRAX also serves as a reliable research guide; specifically, the two first articles build upon this well-established research stream. (Bahmani et al., 2022; Brocke et al., 2013; Fridgen et al., 2022; Fridgen et al., 2018; Halbrügge et al., 2021; Leinauer et al., 2022; Miletić et al., 2021; Miletić et al., 2022; Wederhake et al., 2022). Furthermore, this dissertation draws on the work on decentralization from the Information Management research group of the University of Zurich (Ciriello et al., 2018; Zavolokina et al., 2016; Zavolokina et al., 2017; Zavolokina et al., 2020)

Moreover, this dissertation has been financially supported by the European Union (EU) within its Horizon 2020 programme, project MDOT (Medical Device Obligations Taskforce), Grant agreement 814654. Additionally, by PayPal and the Luxembourg National Research Fund FNR (P17/IS/13342933/PayPal-FNR/Chair in DFS/Gilbert Fridgen).

Declaration of generative AI and AI-assisted technologies in the writing process

I, Joaquin Delgado Fernandez, declare that this thesis is solely my original work and has not been previously submitted for any other degree or professional qualification. I have acknowledged any contributions made by other authors in jointly-authored research papers and have provided accurate citations and references throughout the thesis.

To exclusively improve the readability, coherence of the content, and fix grammatical mistakes previously present in the text, I have utilized AI tools, including ChatGPT, Deepl Write and Writefull during the preparation of this work. However, I have carefully reviewed and edited the generated content to ensure its accuracy, relevance, and alignment with the intended message of this dissertation. Thus, I assume full responsibility for the final content presented in this dissertation.

Luxembourg, September 26, 2023

Joaquín Delgado
Fernández

VI | References

Abbas, A. E., W. Agahari, M. Van de Ven, A. Zuiderwijk, and M. De Reuver (2021). "Business data sharing through data marketplaces: A systematic literature review". In: *Journal of Theoretical and Applied Electronic Commerce Research* 16.7, pp. 3321–3339.

Abbasi, A., S. Sarker, and R. H. Chiang (2016). "Big data research in information systems: Toward an inclusive research agenda". In: *Journal of the association for information systems* 17.2, p. 3.

Acquisti, A. and R. Gross (2006). "Imagined communities: Awareness, information sharing, and privacy on the Facebook". In: *Privacy Enhancing Technologies: 6th International Workshop, PET 2006, Cambridge, UK, June 28-30, 2006, Revised Selected Papers 6*. Springer, pp. 36–58.

Adadi, A. (2021). "A survey on data-efficient algorithms in big data era". In: *Journal of Big Data* 8. DOI: 10.1186/s40537-021-00419-9.

Akhter, S. H. and F. Robles (2006). "Leveraging internal competency and managing environmental uncertainty: Propensity to collaborate in international markets". In: *International Marketing Review* 23.1, pp. 98–115.

Albrecht, S., S. Reichert, J. Schmid, J. Strüker, D. Neumann, and G. Fridgen (2018). "Dynamics of blockchain implementation-a case study from the energy sector". In.

Altman, E. I. (2002). "Managing Credit Risk: A Challenge for the New Millennium". In: *Economic Notes* 31.2, pp. 201–214. DOI: 10.1111/1468-0300.00084. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0300.00084. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0300.00084.

Andrew, G., O. Thakkar, B. McMahan, and S. Ramaswamy (2021). "Differentially private learning with adaptive clipping". In: *Advances in Neural Information Processing Systems* 34.

Aral, S. and P. Weill (2007). “IT assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation”. In: *Organization science* 18.5, pp. 763–780.

Atrias (2021). *The central hub in providing information in the energy market.* Accessed: 2021-05-28. URL: https://www.atrias.be/.

Attard, J., F. Orlandi, and S. Auer (2016). “Data value networks: Enabling a new data ecosystem”. In: *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, pp. 453–456.

Bagdasaryan, E., A. Veit, Y. Hua, D. Estrin, and V. Shmatikov (2020). “How to backdoor federated learning”. In: *International conference on artificial intelligence and statistics*. PMLR, pp. 2938–2948.

Bahmani, R., C. v. Stiphoudt, S. P. Menci, M. SchÖpf, and G. Fridgen (2022). “Optimal industrial flexibility scheduling based on generic data format”. In: *Energy Informatics* 5.1, p. 26. DOI: 10.1186/s42162-022-00198-4. URL: https://doi.org/10.1186/s42162-022-00198-4.

Balta, D., M. Sellami, P. Kuhn, U. Schöpp, M. Buchinger, N. Baracaldo, A. Anwar, H. Ludwig, M. Sinn, M. Purcell, et al. (2021). “Accountable Federated Machine Learning in Government: Engineering and Management Insights”. In: *International Conference on Electronic Participation*. Springer.

Bansal, A., R. J. Kauffman, and R. R. Weitz (1993). “Comparing the Modeling Performance of Regression and Neural Networks as Data Quality Varies: A Business Value Approach”. In: *Journal of Management Information Systems* 10.1, pp. 11–32. ISSN: 07421222. URL: http://www.jstor.org/stable/40398029.

Barbereau, T., R. Smethurst, O. Papageorgiou, J. Sedlmeir, and G. Fridgen (2023). “Decentralised Finance’s timocratic governance: The distribution and exercise of tokenised voting rights”. In: *Technology in Society* 73, p. 102251. ISSN: 0160-791X. DOI: https://doi.org/10.1016/j.techsoc.2023.102251. URL: https://www.sciencedirect.com/science/article/pii/S0160791X23000568.

Benmalek, M., M. A. Benrekia, and Y. Challal (2022). “Security of federated learning: Attacks, defensive mechanisms, and challenges”. In: *Revue des Sciences et Technologies de l’Information-Série RIA: Revue d’Intelligence Artificielle* 36.1, pp. 49–59.

Bi, X., A. Gupta, and M. Yang (2023). *Understanding Partnership Formation and Repeated Contributions in Federated Learning: An Analytical Investigation*. en. SSRN Scholarly

Paper. Rochester, NY. DOI: 10.2139/ssrn.3986446. URL: https://papers.ssrn.com/abstract=3986446 (visited on 07/26/2023).

Bollenbach, J., S. Neubig, A. Hein, R. Keller, and H. Krcmar (2022). "Using machine learning to predict POI occupancy to reduce overcrowding". In.

Bonawitz, K., V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth (2017). "Practical secure aggregation for privacy-preserving machine learning". In: *ACM Conference on Computer & Communications Security*.

Borgman, C. L. (2012). "The conundrum of sharing research data". In: *Journal of the American Society for Information Science and Technology* 63.6, pp. 1059–1078. DOI: 10.1002/asi.22634. eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.22634. URL: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.22634.

Bouncken, R. B., J. Gast, S. Kraus, and M. Bogers (2015). "Coopetition: a systematic review, synthesis, and future research directions". In: *Review of Managerial Science* 9, pp. 577–601.

Brocke, J. vom, G. Fridgen, H. Hasan, W. Ketter, and R. Watson (2013). "Energy informatics: designing a discipline (and possible lessons for the IS community)". In.

Burkhart, S., A. Unterweger, G. Eibl, and D. Engel (2018). "Detecting swimming pools in 15-minute load data". In: *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, pp. 1651–1655.

Chan, C. M., S. Y. Teoh, A. Yeow, and G. Pan (2019). "Agility in responding to disruptive digital innovation: Case study of an SME". In: *Information Systems Journal* 29.2, pp. 436–455.

Chang, K., N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D. Rubin, and J. Kalpathy-Cramer (2018). "Distributed deep learning networks among institutions for medical imaging". In: *Journal of the American Medical Informatics Association : JAMIA* 25. DOI: 10.1093/jamia/ocy017.

Chen, H., R. H. Chiang, and V. C. Storey (2012). "Business intelligence and analytics: From big data to big impact". In: *MIS Quarterly* 36.4.

Chen, N., B. Ribeiro, and A. Chen (2016). "Financial credit risk assessment: a recent review". In: *Artificial Intelligence Review* 45. DOI: 10.1007/s10462-015-9434-x.

Chen, Y.-H., H.-H. Chen, and P.-C. Huang (2018). “Enhancing the data privacy for public data lakes”. In: *2018 IEEE International Conference on Applied System Invention (ICASI)*. IEEE, pp. 1065–1068.

Ciriello, R. F., A. Richter, and G. Schwabe (2018). “Digital innovation”. In: *Business & Information Systems Engineering* 60, pp. 563–569.

Ciuriak, D. (2019). “Unpacking the Valuation of Data in the Data-Driven Economy”. In: *Available at SSRN 3379133*.

Dinev, T., H. Xu, J. H. Smith, and P. Hart (2013). “Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts”. In: *European Journal of Information Systems* 22.3, pp. 295–316.

Dwork, C. (2006). “Differential Privacy”. In: *Automata, Languages and Programming*. Ed. by M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–12.

Dwork, C. and A. Roth (2014). “The Algorithmic Foundations of Differential Privacy”. In: *Found. Trends Theor. Comput. Sci.* 9.3–4, 211–407. ISSN: 1551-305X. DOI: 10.1561/0400000042. URL: https://doi.org/10.1561/0400000042.

Economist, T. (2017). *The world’s most valuable resource is no longer oil, but data*. URL: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

Eder, J. and V. A. Shekhovtsov (2020). “Data quality for medical data lakelands”. In: *Future Data and Security Engineering: 7th International Conference, FDSE 2020, Quy Nhon, Vietnam, November 25–27, 2020, Proceedings 7*. Springer, pp. 28–43.

Eibl, G., K. Bao, P.-W. Grassal, D. Bernau, and H. Schmeck (2018). “The influence of differential privacy on short term electric load forecasting”. In: *Energy Informatics* 1.1, pp. 93–113.

Eibl, G., S. Burkhart, and D. Engel (2019). “Insights into unsupervised holiday detection from low-resolution smart metering data”. In: *Information Systems Security and Privacy: 4th International Conference, ICISSP 2018, Funchal-Madeira, Portugal, January 22-24, 2018, Revised Selected Papers 4*. Springer, pp. 281–302.

Eibl, G., D. Engel, and C. Neureiter (2015). “Privacy-relevant smart metering use cases”. In: *2015 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, pp. 1387–1392.

Ekbia, H., M. Mattioli, I. Kouper, G. Arave, A. Ghazi, T. Bowman, V. Suri, A. Tsou, S. Weingart, and C. Sugimoto (2015). "Big Data, Bigger Dilemmas: A Critical Review". In: *Journal of the Association for Information Science and Technology* 66. DOI: 10.1002/asi.23294.

Elhub (2021). *Netural data hub fir metering data and market processes.* Accessed: 2021-05-28. URL: https://elhub.no/en/#.

Entwistle, T. (2005). "Why are local authorities reluctant to externalise (and do they have good reason)?" In: *Environment and Planning C: Government and Policy* 23.2, pp. 191–206.

Eurich, M., N. Oertel, and R. Boutellier (2010). "The impact of perceived privacy risks on organizations' willingness to share item-level event data across the supply chain". In: *Electronic Commerce Research* 10, pp. 423–440.

European Commission, Directorate-General for Energy, G Küpper, M Cavarretta, A Ehrenmann, E Naffah, J Szilagyi, D Guldentops, N Rozai, and L Charlier ((2020)). *Format and procedures for electricity (and gas) data access and exchange in Member States*. Publications Office. DOI: doi/10.2833/719689.

Fan, Z., P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, M. Sooriyabandara, Z. Zhu, S. Lambotharan, and W. H. Chin (2013). "Smart grid communications: Overview of research challenges, solutions, and standardization activities". In: *IEEE Communications Surveys and Tutorials* 15 (1), pp. 21–38. ISSN: 1553877X. DOI: 10.1109/SURV.2011.122211.00021.

Fang, H. (2015). "Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem". In: *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE, pp. 820–824.

Fatima, S., K. C. Desouza, and G. S. Dawson (2020). "National strategic artificial intelligence plans: A multi-dimensional analysis". In: *Economic Analysis and Policy* 67, pp. 178–194.

Fekri, M. N., K. Grolinger, and S. Mir (2021). "Distributed load forecasting using smart meter data: Federated learning with Recurrent Neural Networks". In: *International Journal of Electrical Power & Energy Systems*, p. 107669.

Ferner, C., G. Eibl, A. Unterweger, S. Burkhart, and S. Wegenkittl (2019). “Pool Detection from Smart Metering Data with Convolutional Neural Networks”. In: *Energy Informatics* 2 (1), pp. 1–9. DOI: https://doi.org/10.1186/s42162-019-0097-8.

Fernández, J. D., T. Barbereau, and O. Papageorgiou (2022a). *Agent-based Model of Initial Token Allocations: Evaluating Wealth Concentration in Fair Launches*. arXiv: 2208.10271.

Fernández, J. D., M. Brennecke, T. Barbereau, A. Rieger, and G. Fridgen (2023a). *Federated Learning: Organizational Opportunities, Challenges, and Adoption Strategies*. arXiv: 2308.02219.

Fernández, J. D., S. P. Menci, C. M. Lee, A. Rieger, and G. Fridgen (2022b). “Privacy-preserving federated learning for residential short-term load forecasting”. In: *Applied Energy* 326, p. 119915. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2022.119915. URL: https://www.sciencedirect.com/science/article/pii/S0306261922011722.

Fernández, J. D., S. P. Menci, and I. Pavic (2023b). “Towards a peer-to-peer residential short-term load forecasting with federated learning”. In: pp. 1–6. DOI: 10.1109/PowerTech55446.2023.10202782.

Fernández, J. D., L. Willburger, C. Wiethe, S. Wenninger, and G. Fridgen (n.d.). “Scaling Smart Cities with Federated Learning–Balancing Accuracy and Privacy for Building Energy Performance Prediction”. In: *Available at SSRN 4489420* ().

Freddie Mac (2021). *Single Family Loan-Level Dataset*. Accessed: 2021-08-11. URL: http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.page.

Fredrikson, M., S. Jha, and T. Ristenpart (2015). “Model inversion attacks that exploit confidence information and basic countermeasures”. In: *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pp. 1322–1333.

Fridgen, G., S. Halbrügge, M.-F. Körner, A. Michaelis, and M. Weibelzahl (2022). “Artificial Intelligence in Energy Demand Response: A Taxonomy of Input Data Requirements”. In.

Fridgen, G., M. Kahlen, W. Ketter, A. Rieger, and M. Thimmel (2018). “One rate does not fit all: An empirical analysis of electricity tariffs for residential microgrids”. In: *Applied energy* 210, pp. 800–814.

Galindo, J. and P. Tamayo (2000). “Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications”. In: *Computational Economics* 15, pp. 107–43. DOI: 10.1023/A:1008699112516.

Geiping, J., H. Bauermeister, H. Dröge, and M. Moeller (2020). "Inverting gradients-how easy is it to break privacy in federated learning?" In: *Advances in Neural Information Processing Systems* 33, pp. 16937–16947.

Giebler, C., C. Gröger, E. Hoos, H. Schwarz, and B. Mitschang (2019). "Leveraging the data lake: Current state and challenges". In: *Big Data Analytics and Knowledge Discovery: 21st International Conference, DaWaK 2019, Linz, Austria, August 26–29, 2019, Proceedings 21*. Springer, pp. 179–188.

Goyal, S. and S. Joshi (2003). "Networks of collaboration in oligopoly". In: *Games and Economic behavior* 43.1, pp. 57–85.

Gupta, K. (2019). "Artificial intelligence for governance in India: Prioritizing the challenges using analytic hierarchy process (AHP)". In: *Int. J. Recent Technol. Eng* 8, pp. 3756–3762.

Halbrügge, S., P. Schott, M. Weibelzahl, H. U. Buhl, G. Fridgen, and M. Schöpf (2021). "How did the German and other European electricity systems react to the COVID-19 pandemic?" In: *Applied Energy* 285, p. 116370.

Han, F., T. Pu, M. Li, and G. Taylor (2020). "Short-term forecasting of individual residential load based on deep learning and K-means clustering". In: *CSEE Journal of Power and Energy Systems* 7.2, pp. 261–269.

Haney, A., T. Jamasb, and M. Pollitt (2009). "Smart Metering and Electricity Demand: Technology, Economics and International Experience". In: *Faculty of Economics, University of Cambridge, Cambridge Working Papers in Economics*.

He, Y., F. Luo, G. Ranzi, and W. Kong (2021). "Short-Term Residential Load Forecasting Based on Federated Learning and Load Clustering". In: *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 77–82.

Heitfield, E. (2009). "Parameter Uncertainty and the Credit Risk of Collateralized Debt Obligations". In: *Risk Management*.

Hippert, H., C. Pedreira, and R. Souza (2001). "Neural networks for short-term load forecasting: a review and evaluation". In: *IEEE Transactions on Power Systems* 16.1, pp. 44–55. DOI: 10.1109/59.910780.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Hoffmann, W., D. Lavie, J. J. Reuer, and A. Shipilov (2018). “The interplay of competition and cooperation”. In: *Strategic Management Journal* 39.12, pp. 3033–3052.

Hornek, T., S. P. Menci, J. D. Fernández, and I. Pavic (2023). “Comparative Analysis of Baseline Models for Rolling Price Forecasts in the German Continuous Intraday Electricity Market”. In.

Houellebecq, M. and F. Wynne (2001). *The elementary particles*. Vintage International.

Hsu, J., M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth (2014). “Differential privacy: An economic method for choosing epsilon”. In: *2014 IEEE 27th Computer Security Foundations Symposium*. IEEE, pp. 398–410.

Isaak, J. and M. J. Hanna (2018). “User data privacy: Facebook, Cambridge Analytica, and privacy protection”. In: *Computer*.

Janiesch, C., P. Zschech, and K. Heinrich (2021). “Machine learning and deep learning”. In: *Electronic Markets*, pp. 1–11.

Jha, S., M. Guillen, and J. Westland (2012). “Employing transaction aggregation strategy to detect credit card fraud”. In: *Expert Systems with Applications* 39, 12650–12657. DOI: 10.1016/j.eswa.2012.05.018.

Jordan, M. I. and T. M. Mitchell (2015). “Machine learning: Trends, perspectives, and prospects”. In: *Science*.

Kairouz, P. et al. (2021).

Kaissis, G., A. Ziller, J. Passerat-Palmbach, T. Ryffel, D. Usynin, A. Trask, I. Lima, J. Mancuso, F. Jungmann, M.-M. Steinborn, A. Saleh, M. Makowski, D. Rueckert, and R. Braren (2021). “End-to-end privacy preserving deep learning on multi-institutional medical imaging”. In: *Nature Machine Intelligence*, pp. 1–12. DOI: 10.1038/s42256-021-00337-8.

Kalra, S., J. Wen, J. C. Cresswell, M. Volkovs, and H. R. Tizhoosh (2023). “Decentralized federated learning through proxy model sharing”. In: *Nature Communications* 14.1, p. 2899. DOI: 10.1038/s41467-023-38569-4. URL: https://doi.org/10.1038/s41467-023-38569-4.

Kang, J., Z. Xiong, D. Niyato, S. Xie, and J. Zhang (2019). “Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory”. In: *IEEE Internet of Things Journal*, pp. 1–5.

Khan, Z. A., T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik (2020). “Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework". In: *Sensors* 20.5. ISSN: 1424-8220. DOI: 10.3390/s20051399. URL: https://www.mdpi.com/1424-8220/20/5/1399.

Kim, H, M Marwah, M. F. Arlitt, G Lyon, and J Han (2011). "Unsupervised Disaggregation of Low Frequency Power Measurements". In: pp. 747–758.

Kohli, N. and P. Laskowski (2018). "Epsilon voting: Mechanism design for parameter selection in differential privacy". In: *2018 IEEE Symposium on Privacy-Aware Computing (PAC)*. IEEE, pp. 19–30.

Kolter, J. Z. and T. Jaakkola (2012). "Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation". In: *Journal of Machine Learning Research - Proceedings Track* 22, pp. 1472–1482.

Konečný, J., H. B. McMahan, D. Ramage, and P. Richtárik (2016). *Federated Optimization: Distributed Machine Learning for On-Device Intelligence*. DOI: 10.48550/ARXIV.1610.05492. arXiv: 1610.02527 [cs.LG].

Lamport, L., R. E. Shostak, and M. C. Pease (1982). "The Byzantine Generals Problem". In: *ACM Trans. Program. Lang. Syst.* 4, pp. 382–401.

Lee, C. M., J. D. Fernández, S. P. Menci, A. Rieger, and G. Fridgen (2023). "Federated Learning for Credit Risk Assessment". In: *Proceedings of the 56th Hawaii International Conference on System Sciences*, p. 10. DOI: https://hdl.handle.net/10125/102676.

Lee, J. and C. Clifton (2011). "How much is enough? choosing ε for differential privacy". In: *Information Security: 14th International Conference, ISC 2011, Xi'an, China, October 26-29, 2011. Proceedings 14*. Springer, pp. 325–340.

Leinauer, C., P. Schott, G. Fridgen, R. Keller, P. Ollig, and M. Weibelzahl (2022). "Obstacles to demand response: Why industrial companies do not adapt their power consumption to volatile power generation". In: *Energy Policy* 165, p. 112876. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2022.112876. URL: https://www.sciencedirect.com/science/article/pii/S030142152200101X.

Leiponen, A. E. (2002). "Why Do Firms Not Collaborate? The Role of Competencies and Technological Regimes". In.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Hung Byers, et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Marino, D. L., K. Amarasinghe, and M. Manic (2016). “Building energy load forecasting using deep neural networks”. In: *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*. IEEE, pp. 7046–7051.

McAfee, A., E. Brynjolfsson, T. H. Davenport, D. Patil, and D. Barton (2012). “Big data: the management revolution”. In: *Harvard business review* 90.10, pp. 60–68.

McLoughlin, F., A. Duffy, and M. Conlon (2015). “A clustering approach to domestic electricity load profile characterisation using smart metering data”. In: *Applied energy* 141, pp. 190–199.

McMahan, H. B., E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas (2017). “Communication-Efficient Learning of Deep Networks from Decentralized Data”. In: *Artificial Intelligence and Statistics*. PMLR, pp. 1273–1282.

McMahan, H. B., D. Ramage, K. Talwar, and L. Zhang (2018). *Learning Differentially Private Recurrent Language Models*. arXiv: 1710.06963 `[cs.LG]`.

Merendino, A., S. Dibb, M. Meadows, L. Quinn, D. Wilson, L. Simkin, and A. Canhoto (2018). “Big data, big decisions: The impact of big data on board level decision-making”. In: *Journal of Business Research* 93, pp. 67–78.

Mikalef, P. and M. Gupta (2021). “Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance”. In: *Information & Management* 58.3, p. 103434.

Mikalef, P., I. O. Pappas, J. Krogstie, and M. Giannakos (2018). “Big data analytics capabilities: a systematic literature review and research agenda”. In: *Information Systems and e-Business Management* 16, pp. 547–578.

Miletić, M., M Gržanić, I. Pavić, H Pandžić, and T. Capuder (2021). “Dynamic electricity pricing tariffs: Trade-offs for suppliers and consumers”. In.

Miletić, M., I. Pavić, H. Pandžić, and T. Capuder (2022). “Day-ahead Electricity Price Forecasting Using LSTM Networks”. In: *2022 7th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, pp. 1–6.

Muñoz, A., E. F. Sánchez-Úbeda, A. Cruz, and J. Marín (2010). “Short-term Forecasting in Power Systems: A Guided Tour”. In: *Handbook of Power Systems II*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 129–160. ISBN: 978-3-642-12686-4. DOI: 10.1007/978-3-642-12686-4_5. URL: https://doi.org/10.1007/978-3-642-12686-4_5.

Muscio, A. (2007). “The impact of absorptive capacity on SMEs’ collaboration”. In: *Economics of Innovation and New Technology* 16.8, pp. 653–668.

Nalebuff, B. J., A. Brandenburger, and A. Maulana (1996). *Co-opetition*. HarperCollins-Business London.

Nassif, A. B., B. Soudan, M. Azzeh, I. Attilli, and O. AlMulla (2021). “Artificial Intelligence and Statistical Techniques in Short-Term Load Forecasting: A Review”. In: *ArXiv* abs/2201.00437.

Nti, I. K., M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya (2020). “Electricity load forecasting: a systematic review”. In: *Journal of Electrical Systems and Information Technology* 7.1, pp. 1–19.

OECD (2019). *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies*.

Ojo, A. and J. Millard (2017). *Government 3.0–Next Generation Government Technology Infrastructure and Services: Roadmaps, Enabling Technologies & Challenges*. Springer.

Olson, M. (1965). “The logic of collective action Harvard University Press”. In: *Cambridge, MA*.

Osarenkhoe, A. (2010). “A coopetition strategy–a study of inter-firm dynamics between competition and cooperation”. In: *Business Strategy Series* 11.6, pp. 343–362.

Passerat-Palmbach, J., T. Farnan, M. McCoy, J. D. Harris, S. T. Manion, H. L. Flannery, and B. Gleim (2020). “Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data”. In: IEEE.

Pathak, N., D. Lachut, N. Roy, N. Banerjee, and R. Robucci (2018). “Non-Intrusive Air Leakage Detection in Residential Homes”. In: *Proceedings of the 19th International Conference on Distributed Computing and Networking*, pp. 1–10.

Radovanovic, D., A. Unterweger, G. Eibl, D. Engel, and J. Reichl (2022). “How unique is weekly smart meter data?” In: *Energy Informatics* 5.1, pp. 1–10.

Rückel, T., J. Sedlmeir, and P. Hofmann (2022). “Fairness, integrity, and privacy in a scalable blockchain-based federated learning system”. In: *Computer Networks* 202, p. 108621. ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2021.108621. URL: https://www.sciencedirect.com/science/article/pii/S1389128621005132.

Sah, M. P. and A. Singh (2022). “Aggregation Techniques in Federated Learning: Comprehensive Survey, Challenges and Opportunities”. In: *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, pp. 1962–1967.

Savi, M. and F. Olivadese (2021). “Short-Term Energy Consumption Forecasting at the Edge: A Federated Learning Approach”. In: *IEEE Access* 9, pp. 95949–95969. DOI: 10.1109/ACCESS.2021.3094089.

Saxena, A., M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, et al. (2017). “A review of clustering techniques and developments”. In: *Neurocomputing* 267, pp. 664–681. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2017.06.053. URL: https://www.sciencedirect.com/science/article/pii/S0925231217311815.

Schlatt, V., J. Sedlmeir, J. Traue, and F. Völter (2023). “Harmonizing Sensitive Data Exchange and Double-Spending Prevention Through Blockchain and Digital Wallets: The Case of E-Prescription Management”. In: *Distrib. Ledger Technol.* 2.1. ISSN: 2769-6472. DOI: 10.1145/3571509. URL: https://doi.org/10.1145/3571509.

Sedlmeir, J., H. U. Buhl, G. Fridgen, and R. Keller (2020). “The energy consumption of blockchain technology: Beyond myth”. In: *Business & Information Systems Engineering* 62.6, pp. 599–608.

Sehovac, L. and K. Grolinger (2020). “Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks With Attention”. In: *IEEE Access* 8, pp. 36411–36426. DOI: 10.1109/ACCESS.2020.2975738.

Shakespeare, W. (1598). *The merchant of Venice*.

Shamir, A. (1979). “How to share a secret”. In: *Communications of the ACM* 22.11, pp. 612–613.

Shejwalkar, V. and A. Houmansadr (2021). “Manipulating the byzantine: Optimizing model poisoning attacks and defenses for federated learning”. In: *NDSS*.

Shen, T., J. Zhang, X. Jia, F. Zhang, G. Huang, P. Zhou, K. Kuang, F. Wu, and C. Wu (2020). *Federated Mutual Learning*. arXiv: 2006.16765 [cs.LG].

Shmueli, G. and O. R. Koppius (2011). “Predictive analytics in information systems research”. In: *MIS quarterly*, pp. 553–572.

Shrader, R. C. (2001). “Collaboration and performance in foreign markets: The case of young high-technology manufacturing firms”. In: *Academy of Management journal* 44.1, pp. 45–60.

Sprenkamp, K., J. D. Fernández, S. Eckhardt, and L. Zavolokina (2023). “Federated Learning as a Solution for Problems Related to Intergovernmental Data Sharing”. In: *Proceedings of the 56th Hawaii International Conference on System Sciences*, p. 10. DOI: https://hdl.handle.net/10125/102838.

Sun, T. Q. and R. Medaglia (2019). “Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare”. In: *Government Information Quarterly* 36.2, pp. 368–383.

Syed, D., H. Abu-Rub, A. Ghrayeb, S. S. Refaat, M. Houchati, O. Bouhali, and S. Bañales (2021). “Deep Learning-Based Short-Term Load Forecasting Approach in Smart Grid With Clustering and Consumption Pattern Recognition”. In: *IEEE Access* 9, pp. 54992–55008.

Taleb, N. N. (2020). “Statistical consequences of fat tails: Real world preasymptotics, epistemology, and applications”. In: *arXiv preprint arXiv:2001.10488*.

The European Commission (2009). “Directive 2009/72/EC of the European Parliament and of the council of 13 July 2009 concerning common rules for the internal market in electricity and repealing Directive 003/54/EC”. In: *Official Journal of the Europan Union* L 211/55.

Thiebes, S., S. Lins, and A. Sunyaev (2021). “Trustworthy artificial intelligence”. In: *Electronic Markets* 31.2, pp. 447–464.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

Voskob, M. and N. Punin (2003). “Data mining and Privacy in Public Sector using Intelligent Agents (discussion paper)”. In: *arXiv preprint cs/0311050*.

Walczak, S. (2001). “An empirical analysis of data requirements for financial forecasting with neural networks”. In: *Journal of management information systems* 17.4, pp. 203–222.

Wang, Y. (2020). *Smart meter data analytics : electricity consumer behavior modeling, aggregation, and forecasting /*. eng. Beijing ; Science Press. ISBN: 9789811526244.

Wederhake, L., S. Wenninger, C. Wiethe, and G. Fridgen (2022). “On the surplus accuracy of data-driven energy quantification methods in the residential sector”. In: *Energy Informatics* 5.1, p. 7. DOI: 10.1186/s42162-022-00194-8. URL: https://doi.org/10.1186/s42162-022-00194-8.

Winter, S., N. Berente, J. Howison, and B. Butler (2014). “Beyond the organizational ‘container’: Conceptualizing 21st century sociotechnical work”. In: *Information and Organization* 24.4, pp. 250–269.

Wirtz, B. W., J. C. Weyerer, and B. J. Sturm (2020). “The dark sides of artificial intelligence: An integrated AI governance framework for public administration”. In: *International Journal of Public Administration* 43.9, pp. 818–829.

Wood, A., M. Altman, A. Bembenek, M. Bun, M. Gaboardi, J. Honaker, K. Nissim, D. R. OBrien, T. Steinke, and S. Vadhan (2018). "Differential privacy: A primer for a non-technical audience". In: *Vanderbilt Journal of Entertainment & Technology Law* 21.1, pp. 209–275. URL: http://www.jetlaw.org/journal-archives/volume-21/volume-21-issue-1/differential-privacy-a-primer-for-a-non-technical-audience/.

Xu, R., N. Baracaldo, Y. Zhou, A. Anwar, S. Kadhe, and H. Ludwig (2022). "DeTrust-FL: Privacy-Preserving Federated Learning in Decentralized Trust Setting". In: *2022 IEEE 15th International Conference on Cloud Computing (CLOUD)*. IEEE, pp. 417–426.

Yan, K., X. Wang, Y. Du, N. Jin, H. Huang, and H. Zhou (2018). "Multi-Step Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy". In: *Energies* 11.11. ISSN: 1996-1073. DOI: 10.3390/en11113089. URL: https://www.mdpi.com/1996-1073/11/11/3089.

Yu, H., Z. Liu, Y. Liu, T. Chen, M. Cong, X. Weng, D. Niyato, and Q. Yang (2020). "A Sustainable Incentive Scheme for Federated Learning". In: *IEEE Intelligent Systems* 35.4, pp. 58–69. DOI: 10.1109/MIS.2020.2987774.

Zavolokina, L., M. Dolata, and G. Schwabe (2016). "The FinTech phenomenon: antecedents of financial innovation perceived by the popular press". In: *Financial Innovation* 2.1, pp. 1–16.

Zavolokina, L., M. Dolata, and G. Schwabe (2017). "FinTech transformation: How IT-enabled innovations shape the financial sector". In: *Enterprise Applications, Markets and Services in the Finance Industry: 8th International Workshop, FinanceCom 2016, Frankfurt, Germany, December 8, 2016, Revised Papers 8*. Springer, pp. 75–88.

Zavolokina, L., N. Zani, and G. Schwabe (2020). "Designing for trust in blockchain platforms". In: *IEEE Transactions on Engineering Management*.

Zhan, Y., P. Li, Z. Qu, D. Zeng, and S. Guo (2020). "A learning-based incentive mechanism for federated learning". In: *IEEE Internet of Things Journal* 7.7, pp. 6360–6368.

Zhang, C., Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao (2021). "A survey on federated learning". In: *Knowledge-Based Systems* 216, p. 106775.

Zuiderwijk, A., Y.-C. Chen, and F. Salem (2021). "Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda". In: *Government Information Quarterly*.

Appendix

A. Publication Overview

This section describes the portfolio of publications from this dissertation. This compromises the publications included in the dissertation as in A.1.. Furthermore, the individual contribution of each of the papers is also disclosed in B.. The full text of the included articles is attached in Appendix C.

A.1. Included publications

The following research papers (RP) included in the dissertation follow journal the percentile on the 2022 Scopus ranking [1], while the conferences follow the 2021 GII-GRIN-SCIE (GGS) [2].

- (Fernández et al., 2022b RP 1): J. D. Fernández, S. P. Menci, C. M. Lee, A. Rieger, and G. Fridgen (2022b). "Privacy-preserving federated learning for residential short-term load forecasting". In: *Applied Energy* 326, p. 119915. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2022.119915. URL: https://www.sciencedirect.com/science/article/pii/S0306261922011722.

  Scopus: **99%**.

- (Fernández et al., 2023b RP 2): J. D. Fernández, S. P. Menci, and I. Pavic (2023b). "Towards a peer-to-peer residential short-term load forecasting with federated learning". In: pp. 1–6. DOI: 10.1109/PowerTech55446.2023.10202782.

  GGS: **B**.

[1] See www.scopus.com

[2] See scie.lcc.uma.es

- (Lee et al., 2023 RP 3): C. M. Lee, J. D. Fernández, S. P. Menci, A. Rieger, and G. Fridgen (2023). "Federated Learning for Credit Risk Assessment". In: *Proceedings of the 56th Hawaii International Conference on System Sciences*, p. 10. DOI: https://hdl.handle.net/10125/102676.

  GGS: **A**.

- (Sprenkamp et al., 2023 RP 4): K. Sprenkamp, J. D. Fernández, S. Eckhardt, and L. Zavolokina (2023). "Federated Learning as a Solution for Problems Related to Intergovernmental Data Sharing". In: *Proceedings of the 56th Hawaii International Conference on System Sciences*, p. 10. DOI: https://hdl.handle.net/10125/102838.

  GGS: **A**.

- (Fernández et al., 2023a RP 5): J. D. Fernández, M. Brennecke, T. Barbereau, A. Rieger, and G. Fridgen (2023a). *Federated Learning: Organizational Opportunities, Challenges, and Adoption Strategies*. arXiv: 2308.02219.

  Scopus: Under Review / Preprint.

A.2. Other peer-reviewed publications not included in this dissertation

- T. Hornek, S. P. Menci, J. D. Fernández, and I. Pavic (2023). "Comparative Analysis of Baseline Models for Rolling Price Forecasts in the German Continuous Intraday Electricity Market". In.
  - Forthcoming in the 15th International Conference in Applied Energy, 2023.

- J. D. Fernández, L. Willburger, C. Wiethe, S. Wenninger, and G. Fridgen (n.d.). "Scaling Smart Cities with Federated Learning–Balancing Accuracy and Privacy for Building Energy Performance Prediction". In: *Available at SSRN 4489420* ().
  - Under review.

- J. D. Fernández, T. Barbereau, and O. Papageorgiou (2022a). *Agent-based Model of Initial Token Allocations: Evaluating Wealth Concentration in Fair Launches*. arXiv: 2208.10271.
  - Under review.

B. Individual Contributions

The papers on this dissertation and the individual contribution of each of the authors are in CRediT roles:

RP 1: Privacy-preserving federated learning for residential short-term load forecasting:

- **Joaquín Delgado Fernández - lead co-authorship**: Conceptualization, Methodology, Data curation, Writing – original draft, Software, Writing – review & editing, Visualization.
- **Sergio Potenciano Menci - subordinate co-authorship**: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Visualization.
- **Chul Min Lee - subordinate co-authorship**: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.
- **Alexander Rieger - subordinate co-authorship**: Writing – review & editing, Supervision.
- **Gilbert Fridgen - subordinate co-authorship**: Writing – review & editing, Supervision, Funding acquisition.

Specifically, I contributed to the development of the paper's concept and created the initial draft. Additionally, I wrote all the code for the simulations and created the original visualization. Furthermore, I was responsible for extracting and pre-processing the dataset. Finally, I participated in the internal review process of the paper and also the needed edits based on the feedback we received during the journal editing stage.

RP 2: Towards a peer-to-peer residential short-term load forecasting with federated learning:

- **Joaquín Delgado Fernández - lead co-authorship**: Conceptualization, Methodology, Data curation, Writing – original draft, Software, Writing – review & editing, Visualization.

- **Sergio Potenciano Menci - subordinate co-authorship**: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Visualization.
- **Ivan Pavić - subordinate co-authorship**: Writing – review & editing, Supervision.

I specifically contributed to the concept formulation of the paper and wrote the first draft. In addition, I coded all the programming and designed the original visualization. Furthermore, I was in charge of extracting and pre-processing the dataset. Finally, I participated in the internal review process of the article as well as revising based on input gathered during the conference editing stage.

RP 3: Federated Learning for Credit Risk Assessment:

- **Chul Min Lee - equal co-authorship**: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing.
- **Joaquín Delgado Fernández - equal co-authorship**: Conceptualization, Methodology, Data curation, Writing – original draft, Software, Writing – review & editing, Visualization.
- **Sergio Potenciano Menci - subordinate co-authorship**: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.
- **Alexander Rieger - subordinate co-authorship**: Writing – review & editing, Supervision.
- **Gilbert Fridgen - subordinate co-authorship**: Writing – review & editing, Supervision, Funding acquisition.

In particular, I helped develop the paper's concept and wrote the first draft with my co-authors. In addition, while my co-author found and managed to obtain the dataset, I wrote all the code responsible for extracting and preprocessing it. I was also responsible for building entirely the prototype. Finally, I participated in the internal review process of the article and also in the editing based on the feedback that we received during the conference editing stage.

RP 4: Federated Learning as a Solution for Problems Related to Intergovernmental Data Sharing:

- **Kilian Sprenkamp- lead co-authorship**: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Visualization.
- **Joaquín Delgado Fernández- subordinate co-authorship**: Conceptualization, Methodology, Data curation, Writing – original draft, Software, Visualization.
- **Sven Eckhardt- subordinate co-authorship**:Writing – review & editing, Supervision.
- **Liudmila Zavolokina- subordinate co-authorship**: Writing – review & editing, Supervision, Funding acquisition.

I helped develop the concept and write the first text, as well as the simulations we provided for the first submission. Following the review, I assisted with the literature review and oversaw the technical elements of the report. Finally, I participated in the internal review process for the article, as well as in the rewriting based on feedback obtained throughout the conference editing stage.

RP 5: Federated Learning: Organizational Opportunities, Challenges, and Adoption Strategies:

- **Joaquín Delgado Fernández - equal co-authorship**: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Visualization.
- **Martin Brennecke - subordinate co-authorship**: Writing – review & editing, Visualization.
- **Tom Josua Barbereau - equal co-authorship**: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Visualization.
- **Alexander Rieger - subordinate co-authorship**: Writing – review & editing, Supervision.
- **Gilbert Fridgen - subordinate co-authorship**: Writing – review & editing, Supervision, Funding acquisition.

Principally, I contributed to the development of the concept of the paper and prepared the initial draft along with the design of the figures and tables. I was responsible for the analysis of the relevant literature, as well as the analysis of the relevant findings on the adoption of FL. Finally, I participated in the internal review process of the paper and also in the necessary revisions based on the feedback that we received during the journal editing stage.

C. Appended Research Papers

Research Paper 1 – *Privacy-preserving federated learning for residential short-term load forecasting*

Full Title:
Privacy-Preserving Federated Learning for Residential Short-Term Load Forecasting

Authors:
Joaquín Delgado Fernández, Sergio Potenciano Menci, Chul Min Lee, Alexander Rieger, Gilbert Fridgen

Published in: Applied Energy (2022)

Keywords:
Deep Neural Networks, Differential Privacy, Federated Learning, Secure Aggregation, Secure Federated Learning, Short-Term Load Forecasting

DOI:
https://doi.org/10.1016/j.apenergy.2022.119915

Abstract

With high levels of intermittent power generation and dynamic demand patterns, accurate forecasts for residential loads have become essential. Smart meters can play an important role when making these forecasts as they provide detailed load data.

However, using smart meter data for load forecasting is challenging due to data privacy requirements. This paper investigates how these requirements can be addressed through a combination of federated learning and privacy preserving techniques such as differential privacy and secure aggregation. For our analysis, we employ a large set of residential load data and simulate how different federated learning models and privacy preserving techniques affect performance and privacy. Our simulations reveal that combining federated learning and privacy preserving techniques can secure both high forecasting accuracy and near-complete privacy. Specifically, we find that such combinations enable a high level of information sharing while ensuring privacy of both the processed load data and forecasting models. Moreover, we identify and discuss challenges of applying federated learning, differential privacy and secure aggregation for residential short-term load forecasting.

1 Introduction

As the supply from intermittent and difficult-to-forecast renewable power sources increases, load forecasting – and especially residential short-term load forecasting (STLF) - is becoming ever more crucial for the reliability of modern power systems (Nti et al., 2020; Petropoulos et al., 2022). Residential STLF covers forecasting windows from a few minutes to a week ahead (ENTSO-E, 2021; Petropoulos et al., 2022). It plays an important role for many operational processes in the power system, such as planning, operating, and scheduling (Lusis et al., 2017; Muñoz et al., (2010)). For instance, it enables energy providers to identify gaps between supply and demand in their customer portfolios. These gaps typically lead to high imbalance costs and ultimately to higher electricity prices for residential customers (Commission for Regulation of Utilities, 2017; Specht and Madlener, 2019).

Traditionally, residential STLF has relied on aggregated load data and reference load profiles (Lusis et al., 2017; Maltais and Gosselin, 2021; Wang et al., 2019). Yet, aggregation and reference profiles are often ill-suited for power systems with a high share of distributed generation and active demand-side management (Lusis et al., 2017; Maltais and Gosselin, 2021). Moreover, they have become less reliable with residential heating and mobility being increasingly electric (International Energy Agency, 2021; Keramidas et al., 2020) and consumption patterns growing more dynamic, for instance, due to fluctuating levels of remote work (Bielecki et al., 2021). These trends make accurate forecasting of individual residential loads an important priority.

There are various traditional methods for more granular STLF, but most build on limiting linearity assumptions (correlation between values and past values) even though residential load patterns are often highly dynamic (Lusis et al., 2017). Examples include time series models that rely on seasonal autoregressive integrated moving averages (ARIMA) (Kaur and Ahuja, 2017; Lusis et al., 2017), exponential smoothing, or linear transfer functions. Residential STFL is thus increasingly relying on methods that can work with non-linear dependencies, such as many Artificial Intelligence (AI) models (Alfares and Nazeeruddin, 2002; Ardabili et al., 2019; Hippert et al., 2001; Kalimoldayev et al., 2020; Khan et al., 2020).

A core challenge for any of these methods is the availability of granular data (Negnevitsky et al., 2009). In many countries, this 'data scarcity' problem is tackled by pushing for advanced metering infrastructure (AMI), which substantially increases the resolution of residential load data (Rashed Mohassel et al., 2014). STLF methods can make use of this data using either 'centralized' or 'decentralized' approaches. Centralized approaches transfer smart meter data to a central forecasting system. While these forecasting systems promise very accurate results, they face a twofold problem. First, they are subject to substantial privacy challenges because smart meter data are often easily attributable to natural persons. That is, data collected from smart meters can be detailed enough to permit the identification of specific customers (Hinterstocker et al., 2017). The transfer and aggregation of smart meter data is thus typically subject to data privacy regulations such as the European Union's General Data Protection Regulation (GDPR) and its obligations and requirements for processing personal data (Kowarik et al., 2016; McKenna et al., 2012). Second, there are considerable regulatory uncertainties. In particular, it is often unclear how device ownership (who owns the smart meter) and aggregation impact data ownership. Moreover, specific regulations for smart meter data are typically absent (European Commission, Directorate-General for Energy et al., (2020); Haney et al., 2009). These regulatory uncertainties often mean that centralized approaches such as Belgium's Atrias (Atrias, 2021), or Norway's Elhub (Elhub, 2021), which provide so-called *data lakes*, may not be desirable.

Decentralized approaches aim to tackle some of these issues by processing smart meter data locally. A particularly promising of these decentralized approaches is Federated Learning (FL) (Konečný et al., 2016; McMahan et al., 2017). Federated Learning is a machine learning technique that offers a collaboration framework for clients. In a so-called 'federation' clients jointly train and share prediction models instead of training data. Although FL cannot guarantee privacy by itself (Geiping et al., 2020; Zhu et al., 2019), it can be combined with privacy-preserving techniques such as differential privacy (DP) and secure aggregation (SecAgg).

Even though such a combination could substantially benefit residential STLF, academic attention to FL has been limited so far (Biswal et al., 2021; Briggs et al., 2021a; Fekri et al., 2021; He et al., 2021; Husnoo et al., 2022; Khalil et al., 2021; Li et al., 2020a; Lin et al., 2022; Savi and Olivadese, 2021; Shi and Xu, 2022; Taïk and Cherkaoui, 2020; Xu et al., 2021) and the two components have mostly been considered mostly in isolation (Barbosa et al., 2016; Chhachhi and Teng, 2021; Eibl and Engel, 2017b). With this paper, we seek to close several gaps in the literature on FL-based STLF: Firstly, we aim to deepen the understanding of FL-based STLF by examining the effects of clustering based on Pearson correlation and the effects of architectural complexity. Secondly, we analyze the privacy and performance effects of adding privacy-preserving techniques (DP and SecAgg) to FL. Third, we identify key challenges associated with using a combination of FL and privacy-preserving techniques.

To do so, we conduct the following analysis: Initially, we identify promising NN architectures from a review of the recent FL literature. Subsequently, we select the most effective of these architectures and investigate six scenarios using real-world historical data. In a first scenario, we evaluate the performance of the selected architecture in a 'centralized' setting to establish a performance benchmark for the remaining five FL scenarios. In the second scenario, we investigate the performance and computational cost effects of moving from a centralized setting to a FL setting. In a third scenario, we then examine the effects of using correlated training data based on Pearson correlation and socio-economic factors. Correlation is typically avoided in non-federated ML models to increase data variability. Yet, for FL models, correlated data may increase forecasting accuracy (Taïk and Cherkaoui, 2020) and mitigate problems with non-IID (non-independent and non-identically distributed) data. In the fourth scenario, we reflect on the trend to work with ever more complex models and explore the effects of increasing the complexity of the NN's architecture. In scenarios 5 and 6, we study how privacy-preserving techniques affect the training and performance of federated models. Specifically, we investigate the effect of different DP implementations (i.e., clipping techniques) and SecAgg on accuracy, privacy, and computational costs.

The remainder of the paper is structured as follows. Section 2 provides an overview of related work on the use of NNs for STLF, FL, and privacy-preserving techniques. Section 3 covers our evaluation method, including the simulation environment, dataset and evaluation metrics. Section 4 describes our evaluation design. It covers the selection of the baseline NN architecture, the specification of the analyzed differential privacy and secure aggregation techniques, the training process for the federated learning models, and the design of six evaluation scenarios. Section 5 presents the evaluation results for the six scenarios. Finally, section 6 provides a synthesis of our results and points out directions for further research.

2 Related work

2.1 Federated learning

In most fields, AI-based methods have already proven their value. However, their performance is highly dependent on the quantity and quality of available training data. Generally speaking, AI-based methods are typically limited by data fragmentation and isolation – mostly due to competitive pressure and tight regulatory frameworks (related to data privacy and security). To address these challenges, McMahan et al. proposed a new technique, FL (Konečný et al., 2016; McMahan et al., 2017). The main idea of FL is to collaboratively train machine learning models between multiple independent clients without moving or revealing the training data. In other words, FL allows competing participants to leverage each others' datasets without revealing their own individual datasets. In doing so, models trained with FL enable more accurate forecasts than models that were independently trained by each client. To date, there are two canonical training algorithms for FL and four different configurations for the distribution of data and errors.

The two canonical training algorithms are: federated stochastic gradient descent (Fed-SGD) and federated averaging (Fed-Avg) (McMahan et al., 2017). Fed-SGD works by averaging the client's gradients after every pass through a local data batch. More specifically, Fed-SGD clients compute gradients of their 'loss' for a sub-set of their data. The loss is a non-parametric function that penalizes bad predictions and to minimize it, the clients need to move toward the empirical minimum by taking steps in the opposite direction of the gradient. Clients subsequently send their locally computed gradients to a central server. The central server aggregates and averages them - either equally or in a weighted manner - to update the model weights. These updated weights are again sent to the clients and each client trains their local model with the updated weights. Training continues in an iterative manner until a pre-defined number of so called communication rounds have been reached or a common goal is achieved. In Fed-SGD, a communication round represents a full pass through all batches.

In Fed-Avg, the clients send their model weights instead of their gradients.Once the central server has received the weights, it aggregates and averages them to arrive at a new 'consensus' that will be sent back to the clients for the next training round. Unlike Fed-SGD, Fed-Avg does not split the training data into batches, which has two effects: the number of communication rounds is reduced substantially (only once per epoch) and an improvement in forecasting accuracy (Fekri et al., 2021; McMahan et al., 2017). As in Fed-SGD, the training process continues until the pre-defined number of epochs has been reached or a common goal is achieved.

Besides different algorithms, FL applications can also differ in their configurations. These configurations depend on how the data is structured. More specifically, they depend on the configuration of the feature space $\mathcal{X}$, the label space $\mathcal{Y}$, and the space formed by the identifiers $\mathcal{I}$. Different setups of the triplet $(\mathcal{X}, \mathcal{Y}, \mathcal{I})$ can be classified as Horizontal, Vertical, Transfer and Assisted Federated Learning (Yang et al., 2019a). Take for instance two clients i and j.

- Horizontal Federated Learning is when i and j share the same feature space such that $\mathcal{X}_i = \mathcal{X}_j$ but their label spaces $\mathcal{Y}$ are different so that $\mathcal{Y}_i \neq \mathcal{Y}_j$. In our residential STLF example, Horizontal FL would be applicable when the model is to be trained on smart meter data from a range of clients with the same feature set (consumption, weather profile, etc.) and the data is held by different companies.

- Vertical Federated Learning is when $\mathcal{I}_i = \mathcal{I}_j$, but $\mathcal{X}_i \neq \mathcal{X}_j$ and $\mathcal{Y}_i \neq \mathcal{Y}_j$. This would be the case, for instance, when two companies have access to the same client but each of them holds a different feature set regarding the client.

- Federated Transfer Learning happens when $\mathcal{X}_i \neq \mathcal{X}_j$, $\mathcal{Y}_i \neq \mathcal{Y}_j$, $\mathcal{I}_i \neq \mathcal{I}_j, \ \forall \mathcal{D}_i, \mathcal{D}_j, i \neq j$. Federated Transfer Learning can be used, for instance, when two companies have different clients and feature sets but want to nevertheless collaboratively train a model.

- Assisted Learning (AL) is done through collided data between clients. Xian et al. (Xian et al., 2020) define collision as when clients with the same data entries of a dataset $\mathcal{D}$ have different feature spaces $\mathcal{I}_i = \mathcal{I}_j, \mathcal{X}_i \neq \mathcal{X}_i \ \forall \mathcal{D}_i, \mathcal{D}_j, i \neq j$. One client may use the errors of another for their own benefit by increasing their training performance.

Regardless of the chosen algorithm and configuration, FL is vulnerable to moral hazard (Kairouz et al., 2021) or so-called 'soft' attacks on the contextual integrity of the shared data. Moral hazard arises because FL is by nature collaborative (McMahan et al., 2016). Multiple clients must work together to train models iteratively using the respective data at their disposal. If one or several of these clients manipulate the joint training process, it does not work. In effect, federated learning requires trust between the clients involved.

2.2 FL-based short term Load forecasting

Short-term load forecasting is a complex, multivariate time series problem. Its complexity is high because residential load data is often replete with irregularities, missing or inaccurate values, and seasonality. Petropoulus et al. (Petropoulos et al., 2022) provide an in-depth overview of these challenges. Yet, they also point out the increasing importance and momentum that STLF has gained over recent years. STLF is crucial because system operators require it for unit commitment and optimal power flow calculations (Li, 2020; Muñoz et al., (2010); Petropoulos et al., 2022). Moreover, it enables utilities, energy suppliers, and distribution grid operators (DSOs) to optimize their customer portfolios, design tariffs, and strategically adapt flexibility offerings (Muñoz et al., (2010); Petropoulos et al., 2022).

STLF typically build on three groups of methods: traditional methods, AI-based methods, and hybrid methods that integrate traditional and AI-based components (Petropoulos et al., 2022). Traditional methods such as ARIMA can capture seasonal trends but fall short when it comes to non-linear patterns and non-aggregated data. At the same time, they are simple to use and have light computational costs (Petropoulos et al., 2022). AI-based methods, in turn, are well suited to identifying non-linear patterns and work well with individual (i.e., residential level) and aggregated data (i.e., substation level) (Lusis et al., 2017; Vos et al., 2018).

Within the larger group of AI-based methods, FL is a relatively new but increasingly popular method for STLF. Our following overview of these FL studies which follows is based on a search in Semantic Scholar using the following search terms: *short-term load forecasting neural networks* and *Federated Learning for Residential Short Term Load Forecasting*.

The first group of studies employ Fed-SGD (He et al., 2021; Lin et al., 2022). He et al. (He et al., 2021) additionally use k-means clustering and compare performance between six scenarios with a different number of clusters in each scenario. Their results suggest that grouping data based on comparable load patterns substantially improves the performance of FL models. Lin et al. (Lin et al., 2022), in turn, focus on limiting the high computational cost of Fed-SGD. To this end, they introduce an asynchronous stochastic gradient descent algorithm with delay computation (ASGD-DC). Specifically, their algorithm uses a Taylor expansion to compensate for the delay of clients with lower computational power.

The second and substantially larger group of studies employ Fed-Avg. Similar to He et al. (He et al., 2021), Briggs et al. (Briggs et al., 2021a), Savi et al. (Savi and Olivadese, 2021), Afaf et al. (Taïk and Cherkaoui, 2020), and Biswal et al. (Biswal et al., 2021) investigate different forms of clustering for Fed-Avg. Their findings suggests that clustering based on k-means and socio-economic factors can also substantially improve the performance of Fed-Avg. With certain caveats, their findings also suggest that its possible to train good models with a small number of clients. Li et al. (Li et al., 2020a), in turn, use Fed-Avg to compare the effects of different federation sizes, ranging the number of clients from 2, to 4, and 6. They also vary the number of training rounds (epochs) from 5 to 15. Their results suggest performance is increased by increasing the number of clients and training rounds.

Xu et al. (Xu et al., 2021) as well as Husnoo et al. (Husnoo et al., 2022) investigate the effect of increasing the number of clients participating in the training rounds. Their results show a considerably drop in performance for the higher participation cases. This drop appears to be the result of non-IDD consumption data between the clients.

Khalil et al. in (Khalil et al., 2021) use Fed-Avg to train a FL model for building control, replicating the use of FL for household training. They consider six floors of a seven-story building as clients. They later personalize the global FL model for the 7th floor - not used in the FL training - by running locally five additional rounds (epochs) and not sharing the data with the global model. Their results suggest that even the personalized FL model can help a smart building controller reduce total electricity consumption using FL.

In terms of relative performance, Fekri et al. (Fekri et al., 2021) find that Fed-Avg provides more accurate results for STLF than Fed-SGD. Shi et al. (Shi and Xu, 2022),in turn, look beyond canonical FL and use a multiple kernel variant of maximum mean discrepancies (MK-MMD) to fine-tune the central server model (global). They train for several rounds using transfer learning to adapt the global model to specific customers. Their results indicate better performance than a canonical Fed-avg implementation.

The works of (Biswal et al., 2021; Briggs et al., 2021a; Fekri et al., 2021; He et al., 2021; Husnoo et al., 2022; Khalil et al., 2021; Li et al., 2020a; Lin et al., 2022; Savi and Olivadese, 2021; Shi and Xu, 2022; Taïk and Cherkaoui, 2020; Xu et al., 2021) provide important stepping-stones in FL-based STLF. In particular, they clearly indicate the prospect of using collaborative training to create accurate forecasting models. However, they provide only limited insights into the challenges of using FL. In particular, it is not yet clear if different but simpler clustering techniques such as Pearson correlation are also effective. Also, prior literature has not yet looked at the effect of architectural complexity. Moreover, existing studies do not or only in a very limited way account for matters of privacy. Thus, this paper aims to provide a better understanding of clustering and architectural complexity and explores the addition of different privacy preserving techniques.

2.3 NN architectures for FL-based short term Load forecasting

The studies presented on FL-based STLF use a range of different NN architectures (Table 1). Overall, the architectures have become deeper (i.e., multi-layered) over time as depth is typically associated with more accurate results (Vos et al., 2018). In terms of layer design, we found Fully Connected layers (FCL), Long Short-term Memory (LSTM) Layers (Hochreiter and Schmidhuber, 1997) and Convolutional Neural Networks (CNN). LSTMs have feedback connections which understand the dependence between items in a sequence and which make them suitable for temporal pattern recognition. CNN layers emulate human retinas and can capture the spatial distribution of graphic patterns. Moreover, we found Encoder-Decoder or autoencoder architectures (Marino et al., 2016). In these architectures, the NN is provided with a sequence (a vector) as an input and maps this sequence to another sequence. Encoder-Decoder architectures reduce the effects of outliers because they transpose the original input space into a differently encoded space (Cho et al., 2014; Sutskever et al., 2014). Sehovac et al. (Sehovac and Grolinger, 2020) present a particular interesting example of a Seq2Seq architecture that includes an attention mechanism to help the decoder extract additional information.

Aside from different layer designs, we also identified hybrid designs. For instance, Kim et al. (Kim and Cho, 2019b) use CNN with LSTM layers to find both spatial and temporal patterns. Building on their work, Tuong et al. (Le et al., app9204237) add a bi-directional LSTM layer to identify temporal trends both forward and backwards in time. Similarly, Zulfiqar Ahmad et al. (Khan et al., 2020) combine Seq2Seq from (Marino et al., 2016) with a CNN layer design. This combination allows for the capture of both temporal and spatial patterns and offers protection against outliers. Shi et al. (Shi et al., 2018) take a different path by clustering and pooling the training data to increase variability and reduce overfitting.

2.4 Privacy preserving techniques for federated learning

Privacy-preserving techniques can support the design of forecasting systems that comply with privacy requirements and regulations (Bennett, 2018; Li et al., 2021; McKenna et al., 2012). From an organizational perspective, these techniques allow competing agents like energy providers to cooperate and integrate with utilities and DSOs (Ben-

Table 1: Neural network architectures for FL-based and non FL-based STLF.

Method	Dataset	Neural Network Architecture	Year
(Marino et al., 2016)	UCI - Individual household electric power consumption	LSTM + Repeat vector + LSTM + 2x FCL	2016
(Kong et al., 2019)	Australia SGDS Smart Grid Dataset	Stacked LSTM + FCL	2017
(Li et al., 2017)	Fremont, CA 15min Retail building electricity load	Missing or incomplete architecture description	2017
(Shi et al., 2018)	Irish CBTs - Residential and SMEs	Stacked LSTM + Pooling mechanism	2018
(Yan et al., 2018)	UK-DALE Domestic Appliance-Level Electricity dataset	2x Conv + 1x LSTM + FCL	2018
(Kim and Cho, 2019a)	UCI - Individual household electric power consumption	Missing or incomplete architecture description	2019
(Kim and Cho, 2019b)	UCI - Individual household electric power consumption	2x Conv + LSTM + 2x FCL	2019
(Le et al., app9204237)	UCI - Individual household electric power consumption	2x Conv + Bi + LSTM + 2x FCL	2019
(Khan et al., 2020)	UCI - Individual household electric power consumption	2x Conv + 2x LSTM (Encoder) + 2x LSTM (Decoder) + 2x FCL	2020
(Taïk and Cherkaoui, 2020)	Pecan Street Research Institute	2x LSTM (same size) + FCL	2020
(Sehovac and Grolinger, 2020)	Non-disclosed or private data	Sequence to Sequence with attention	2020
(Li et al., 2020a)	Global Energy Forecasting Competition 2012	Missing or incomplete architecture description	2020
(Xu et al., 2021)	Pecan Street Research Institute	Missing or incomplete architecture description	2021
(Briggs et al., 2021b)	Low Carbon London Dataset	2x LSTM (same size) + FCL	2021
(He et al., 2021)	Australia SGDS Smart Grid Dataset	2x LSTM (same size) + FCL	2021
(Savi and Olivadese, 2021)	Low Carbon London Dataset	LSTM (64) + LSTM (32) + FCL	2021
(Zhao et al., 2021)	Pecan Street Research Institute	2x LSTM (same size) + FCL	2021
(Biswal et al., 2021)	Commission for Energy Regulation (CER)	Missing or incomplete architecture description	2021
(Khalil et al., 2021)	CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets	Missing or incomplete architecture description	2021
(Shi and Xu, 2022)	Low Carbon London Dataset	Missing or incomplete architecture description	2022
(Lin et al., 2022)	Commission for Energy Regulation (CER)	Missing or incomplete architecture description	2022
(Husnoo et al., 2022)	Solar Home Electricity Data from Eastern Australia	LSTM (256) + LSTM (128) + FCL	2022

nett, 2018; Kowarik et al., 2016). Furthermore, their use might facilitate the creation of local markets that support the energy transition (Pressmair et al., 2021).

Privacy-preserving techniques are especially relevant for FL. Although FL offers considerable improvements over centralized ML methods, it does not guarantee privacy. Firstly, the shared data (gradients or model weights) may allow inadvertent attribution, and secondly, privacy can be compromised through the communication between clients and the central server. For instance, Zhu et al. found a way to use gradient updates to reconstruct the training data of a client (Zhu et al., 2019). This effectively means that gradient updates are to be treated as personal data and that FL requires additional measures when data privacy is required. In the following, we describe two such measures: DP as a way to anonymize training data and SecAgg as a mechanism to enable privacy-sensitive communication between clients and the central server.

Dwork (Dwork, 2006) introduces DP as a technique to guarantee privacy when retrieving information from a dataset. As described in (Dwork and Roth, 2014), "differential privacy addresses the paradox of knowing nothing about an individual while learning useful information about a population." DP hides individual data trends by using additive noise. In more technical terms, Dwork (Dwork, 2006) introduced epsilon differential privacy (ϵ-DP) as follows: *"For every pair of inputs x and y that differ in one row, for every output in S, an adversary should not be able to use the output in S to distinguish between any x and y"*. The privacy budget (ϵ) determines how much of an individual's privacy a query may use, or to what extent it may increase the risk of breaching an individual's privacy. A value of $\epsilon = 0$ represents perfect privacy, which means that privacy cannot be compromised through any analysis on a dataset in question (Wood et al., 2018). Jayaraman et al. (Jayaraman and Evans, 2019) extended the concept of (ϵ-DP) to (ϵ, δ-DP) where δ is the failure probability to better control for the tails of the privacy budget.

DP is typically implemented by adding random noise to data queries. This noise is usually sampled from a Laplacian or Gaussian distribution (Dwork and Roth, 2014). Finding an adequate noise level is crucial but not trivial - especially for FL. Too much noise can not only hide patterns in the data but also complicate convergence of the local models due to the random updates of the patterns during training. Simply speaking, more noise means more privacy, but more noise also means less accuracy.

An alternative to adding noise to the training process or the data is using secure multi-party computation (SMPC) protocols, which enable privacy-preserving communication. One such protocol is SecAgg (Bonawitz et al., 2017). SecAgg uses cryptographic primitives that prevent the central server from reconstructing each client's involvement and contribution. In more technical terms, SecAgg allows a set of distributed, unknown clients to aggregate a value x without revealing the value to the other clients. The backbone of SecAgg is Shamir's *t-out-of-n* Secret Sharing. It enables a user to split a secret s into n shares (Shamir, 1979). To reconstruct the secret, more than $t-1$ shares are needed to retrieve the original secret s. Any allocation with less than $t-1$ shares will provide no information about the original secret. SecAgg implies two main algorithms: sharing and reconstruction. The sharing algorithm transforms a secret into a set of shares of the secret that are each associated with a client. Following (Shamir, 1979), these shares are constructed in such a way that collusion between $t-1$ participants (t being the total number of participants) is insufficient to disclose other clients' private information. The reconstruction algorithm works in the opposite direction. It takes the mentioned shares from the clients and reconstructs the shared secret.

Of the two privacy-preserving techniques, only DP has so far been examined in the context of residential STLF. Chhachhi et all. (Chhachhi and Teng, 2021), Eibl et al. (Eibl and Engel, 2017a), and Zhao et al. (Zhao et al., 2014) use DP to train a 'centralized' machine learning model. More specifically, they perturb the datasets by adding noise drawn from either a Gaussian or Laplacian distribution before each training round of the model. To the best of our knowledge, Zhao et al. (Zhao et al., 2021) are the first to combine FL and DP for STLF. Specifically, they include DP in the training process of a Fed-Avg model. However, they do not systematically analyze different DP parameters. Moreover, they do not look at secure multi-party computation protocols, such as SecAgg.

3 Method

3.1 Simulation environment

The evaluations in this paper are based on simulations we ran on the IRIS Cluster of the high performance computer (HPC) facilities of the University of Luxembourg (Varrette et al., 2014). The simulations ran in an environment with 32 Intel Skylake cores and two NVIDIA Tesla V100 with 16GB or 32GV depending on the allocation. We programmed the federation code in Python and based it on the machine learning framework provided by Tensorflow-Federated [3] (TFF). The DL models are written in Keras (Chollet et al., 2015).

3.2 Dataset

For our simulations, we used a large dataset from the Low Carbon London project, which was conducted by UK Power Networks between November 2011 and February 2014 in London, United Kingdom (herein LCL dataset) (D., 2019). It contains the electrical consumption [kWh] data from 5567 households in a half-an hour resolution. The LCL dataset also contains a socio-technical classification of the households following the ACORN scheme (CACI, 2014) and is divided into individual household entries known as LCLid (Low Carbon London id).

To make the dataset ready for our simulations, we treated it in a 4-step procedure. First, we reduced the resolution of the LCL dataset to hourly values. The down-scaled values in the treated data set are the sum of two subsequent half-hour values in the original data set. This treatment significantly reduced the computational burden of our simulations. Secondly, we trimmed outliers or null values. Thirdly, we scaled all variables to have the same range using a Min-Max scaler. This re-scaling was necessary to ease the FL learning process as all values have to be in a known range, in our case: 0 to 1. Fourthly and finally, we split the dataset into a training and validation dataset. The training dataset (75%) contains electrical consumption data from January to December 2013 and the validation set (25%) covers data from January 2014 to March 2014. In Figure 1, we provide an example of the processed data. It visualizes the electricity consumption [kWh] of 5 randomly selected households for a 2 day period using 1h timestamps.

[3] https://github.com/tensorflow/federated

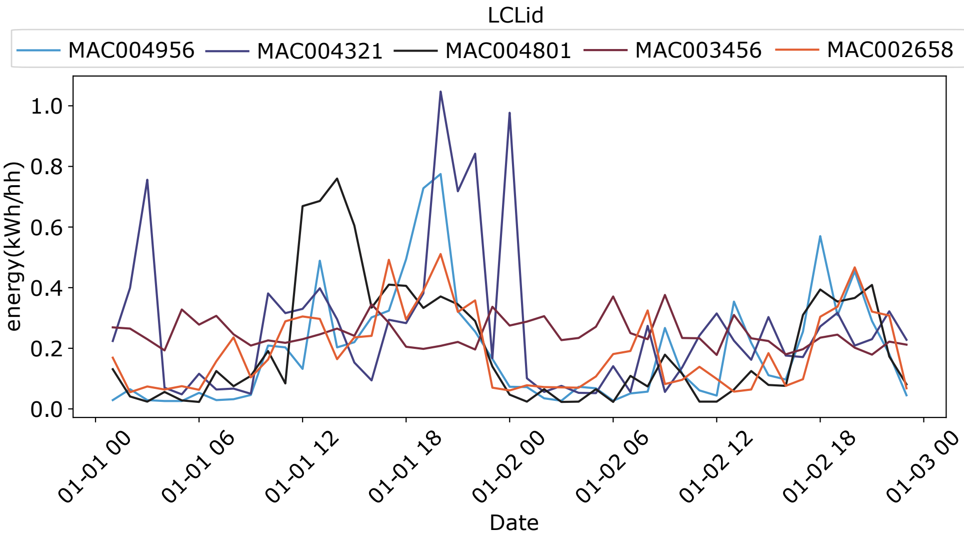


Figure 1: Energy consumption (kWh/h) of 4 LCLIds from 01 January 2013 to 03 January 2013.

3.3 Evaluation metrics

Evaluation metrics offer an important means for the training and testing of forecasting models. However, the use of certain metrics can lead to undesirable results because FL models are known to converge to a *middle point* (Li et al., 2020b). More specifically, FL models optimize the error of prediction with respect to the ground truth. In a distributed environment where there are *many such truths*, the models tend to minimize the mean of the loss across datasets. This tendency can provoke FL models to predict the average of each of the datasets and hence offer promising mean squared errors (MSE, Equation 1) and mean absolute errors (MAE, Equation 2). Such predictions, however, mean that the FL model did not learn local patterns in the data.

Therefore, MSE and MAE are typically not enough to evaluate the performance of a FL model and additional metrics, such as mean absolute percentage error (MAPE, Equation 3) and root mean square error (RMSE, Equation 4), are needed to quantify deviations of model predictions from the ground truths. The formal equations for these four metrics are as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \qquad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \qquad (2)$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| \qquad (3)$$

$$RMSE = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(y_i - x_i)^2} \qquad (4)$$

4 Evaluation

4.1 Selection of a baseline neural network architecture

One crucial aspect for any AI method and specifically FL is the selection of the underlying NN architecture. To pick an architecture for our evaluation, we compared those in Table 1 that had a clear 'implementation guide' we could replicate. For this comparison, we used the metrics described in subsection 3.3, trained the architectures with a maximum of 300 epochs on the training dataset and evaluated them on the evaluation dataset. We used the authors' codes where available and otherwise implemented the architecture ourselves. To limit computational costs, we used an early stopping mechanism for the training, that ended the training when the evaluation metrics did not improve over 10 epochs.

In Figure 2, we illustrate the evaluation results for the twelve architectures we could replicate. Some architectures behaved worse on our dataset than on the dataset used by the respective authors. One possible reason for these differences could be scaling. Kim et al. (Kim and Cho, 2019b; Le et al., app9204237), for instance, worked with a non-scaled dataset. This means that depending on the standard deviation of the dataset σ, the error metrics can differ substantially. For instance, the MSE scales proportionally with the standard deviation: $MSE_{scaled} = MSE_{non-scaled} * \sigma$. To avoid this scaling effect, we calculated all metrics using standardized data (section 5).

Overall, the architectures in (He et al., 2021; Husnoo et al., 2022; Kong et al., 2019; Marino et al., 2016; Savi and Olivadese, 2021; Yan et al., 2018; Zhao et al., 2021) had the lowest MAPE, from 6.7 to 7.1. From these, we selected Marino et al.'s (Marino et al., 2016) autoencoder architecture. Autoencoders are known to perform well even with non-idd data, so we selected the most performant autoencoder architecture among our shortlist of architecures. Marino et al.'s (Marino et al., 2016) architecture uses a 50-neuron encoder layer, a 12-neuron latent space, a 50-neurons decoder layer, and two final layers with 100 and 1 neurons respectively.

For our investigation of the effects of architectural complexity, we selected Khan et al.'s (Khan et al., 2020) architecture as it performed best among the more complex architectures in our sample. Khan et al.'s (Khan et al., 2020) architecture is different from Marino et al.'s (Marino et al., 2016) in that it uses convolutional layers and LSTM.

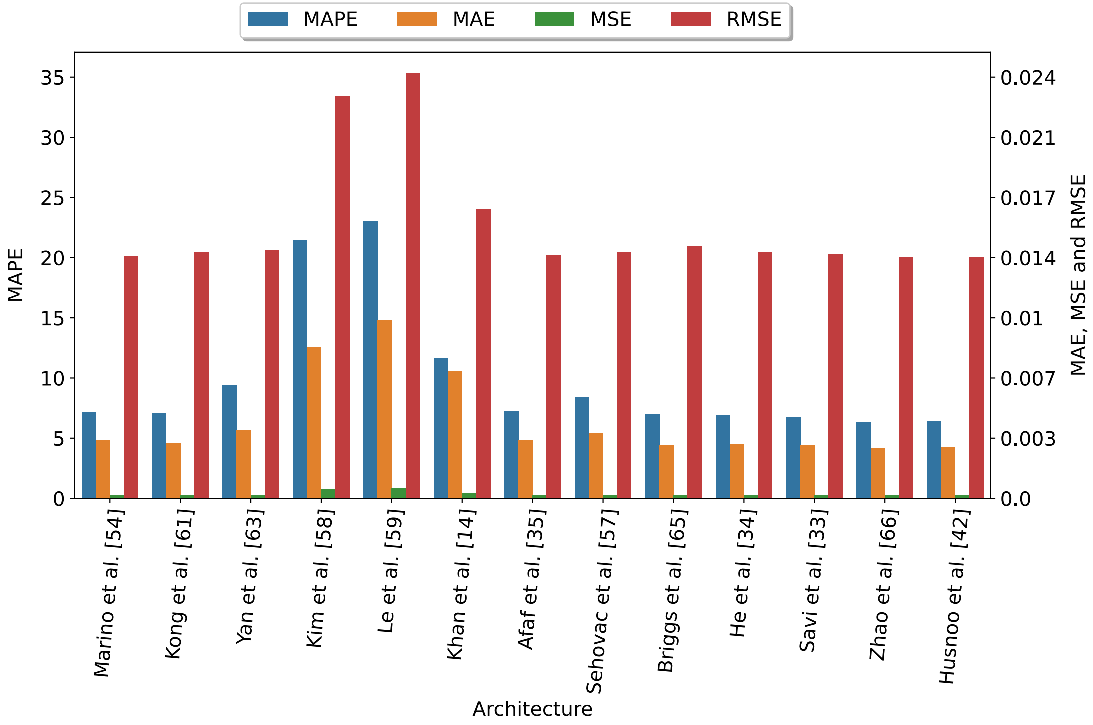


Figure 2: RMSE, MSE, MAE, MAPE of the current literature applied to this paper's dataset.

4.2 FL, differential privacy and secure aggregation set-up

For our simulations, we selected Fed-Avg over Fed-SGD as it requires fewer communication rounds and has better performance (Fekri et al., 2021; Yang et al., 2019b). Moreover, we used a horizontal FL configuration as our clients represent different LCLIds but share the same feature space.

To implement DP, we followed the steps proposed by McMahan et al. (McMahan et al., 2018) rather than those of Chhachhi et al. (Chhachhi and Teng, 2021) and Lu et al. (Lu et al., 2019), in which noise is added to the dataset before the training. McMahan et al. (McMahan et al., 2018) propose the central server to add noise after aggregating the updates of the model weights at every training round (in Fed-Avg). In other words, it differs from canonical Fed-Avg, which aggregates model weights.

The process proposed by McMahan et al. requires the definition of a query function sensitivity ($\mathbb{S}$) and a clipping strategy. The sensitivity of the query function determines the *actuation range* of the added noise. It represents the Euclidean distance between two datasets (C) differing in at most one element k: $\mathbb{S}(\tilde{f}) = max_{C,k} \left\| \tilde{f}(C \cup \{k\}) - \tilde{f}(c) \right\|_2$ (Dwork and Roth, 2014). Considering McMahan et al.'s first lemma (McMahan et al., 2018) and assuming all clients are equally weighted, the sensitivity $\mathbb{S}$ is bounded as $\mathbb{S}(\tilde{f}(c)) \leq S/n$, with n being the number of clients. The vectors in Δ_k include the different model updates computed among the clients.

To bound the sensitivity of the query function, we needed to maintain the models' updates in a known range. One approach to ensure this range control is clipping model updates by a defined value before averaging. There are two strategies to clip the values of a neural network: 'per layer clipping', which applies clipping on a layer basis or 'flat clipping' which applies a clipping value to all the network parameters. Both clipping strategies project the values of the updates into a l2 sphere with the norm determined by the clipping value.

For both, per layer and flat clipping, there are two sub-strategies. One is to clip values using a fixed norm, known as fixed clipping. The second sub-strategy is called adaptive clipping (Andrew et al., 2021). It adapts the clipping norm based on a target quantile (i.e., 0.5) of the data distribution (Andrew et al., 2021).

For the sake of simplicity, we used flat clipping as $\Delta'_k = \pi(\Delta_k, S)$ with S being the overall clipping value for the model updates. At the same time, we implemented both fixed and adaptive flat clipping strategies.

Once we had defined the query sensitivity and applied a flat clipping strategy, we evaluated how noise levels scale with the query sensitivity to obtain the minimum level of noise with a privacy guarantee. We added Gaussian noise as defined by: $N(0, \sigma^2)$ for $\sigma = z \cdot \mathbb{S}$, where z is the noise scale and $\mathbb{S}$ is the sensitivity of the query.

The addition of noise determines the overall privacy protection (ϵ) provided by DP. ϵ varies depending on the amount of noise added and the ratio of clients involved in the training (Q). Q is the ratio of clients selected out of the total which will participate in the next round of training. More noise naturally means more privacy and a lower ϵ. A higher Q, in turn, means less privacy and a higher ϵ (Abadi et al., 2016).

To compute the privacy protection after a query, that is, each training round of our model, we used the privacy accountant provided by Renyi Differential Privacy (RDP) (Mironov, 2017) as it provides a more detailed analysis of the privacy budget than the one created by (McMahan et al., 2018).

For SecAgg, we used the implementation provided by Bonawitz et al. (Bonawitz et al., 2017). Their SecAgg implementation works as a plug-and-play algorithm that does not require any modification. We used SecAgg to ensure privacy-preserving communication between the central server and the clients. By using SecAgg in FL, clients can share their model weights without the central server or another client being able to reconstruct their weights (Shamir, 1979).

4.3 Model operation

In this subsection, we describe how we trained the FL models. For this training, we used 6 steps. We illustrate these steps as well as the additional step that FL-DP requires in Figure 3. FL-SecAgg requires a different additional step, namely the initial sharing of public keys between the clients and central server. Figure 3 does not illustrate this additional public key sharing.

In step 1, the central server initializes the model using Glorot initialization (Glorot and Bengio, 2010). In step two, the central server shares the model with the participating clients. In step three, a subset of clients are selected based on the ratio (Q). Each of these clients in this sub-set then trains the received model on its data. In step four, clients send their model updates to the central server. In step five, the central server averages the aggregated updates and adds noise drawn from a Gaussian distribution in the case of DP (5′ in Figure 3). In step six, the central server returns the averaged updates to the clients. The central server and the clients repeated steps 2 to 6 until they reached 300 epochs.

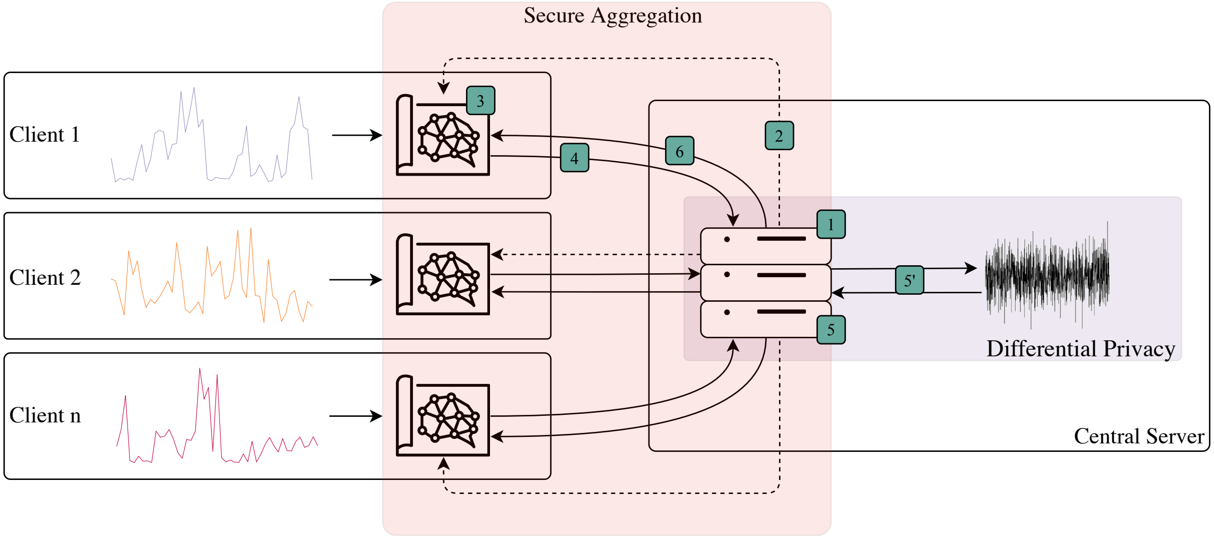


Figure 3: Visual representation of our implementation of Federated Learning with privacy-preserving techniques.

4.4 Scenario design

Overall, we designed a set of six scenarios for our evaluation. Scenario 0 represents a hypothetical scenario in which all clients share their training data with the central server. This 'centralized setting' serves as a benchmark for the other scenarios. In Scenario A, we study the effects of moving from a centralized to a FL setting. In scenario B, we analyze the performance effect of clustering clients based on Pearson correlation. In scenario C, we evaluate the effect of a more complex NN architecture. Lastly, in Scenarios D and E, we study the effects of adding DP and SecAgg to the FL model. We summarize the specifications of the six scenarios in Table 2.

For scenarios 0, A, B, C and E, we ran eight simulations. These simulations evaluate the models' performance with a growing number of clients (federation size). We used the following eight federation sizes: 2, 5, 8, 11, 14, 17, 20, and 23 clients. Each of these clients worked with data from one LCLid. We had to limit the maximum number of clients to 23 to control for computational cost as we simulated all clients and the communication between them in one virtual environment. In effect, every additional client did not add computational power but computational overhead.

We provide an overview of the hyperparameters for scenarios 0, A, B, C and E in Table 2. Table 4 provides the hyperparameters for the DP implementation in Scenario D.

Table 2: Scenarios considered.

Scenario	Privacy-Preserving Technique	NN Architecture	Imposed Correlation
0	-	Marino et al. (Marino et al., 2016)	✗
A	-	Marino et al. (Marino et al., 2016)	✗
B	-	Marino et al. (Marino et al., 2016)	✓
C	-	Khan et al. (Khan et al., 2020)	✗
D	Differential Privacy	He et al. (He et al., 2021)	✗
E	Secure Aggregation	Marino et al. (Marino et al., 2016)	✗

Table 3: Hyperparameters for scenarios A,B,C and E. Those marked with * the ones used in scenario 0.

Parameter	Value
Number of internal rounds before averaging	5
NN architecture	Marino et al. (Marino et al., 2016) * and Khan et al. (Khan et al., 2020)
Ratio of clients involved per round (Q)	1
Total number of clients (w)	Subject to federation size
Optimizer	Adam *
Optimizer learning rate (L_r)	0.01 *
Batch size	256 *
Number of communication rounds	300 *
Number of internal epochs after training	Not applicable

Table 4: Hyperparameters for scenario D.

Parameter	Value
Number of internal rounds before averaging	5
NN Architecture	He et al. (He et al., 2021)
Ratio of clients involved per round (Q)	0.1
Total number of clients (w)	100
Optimizer	Adam
Optimizer learning rate (L_r)	0.01
Batch size	64
Number of communication rounds	100
Number of internal epochs after training	1

5 Evaluation results

5.1 Scenario 0: Centralized setting

Scenario 0 analyzes the performance of a centralized setting, in which the clients send their data to a central server that trains a single model on the aggregated data. Scenario 0 uses the NN architecture presented by Marino et al. (Marino et al., 2016). Similar to the architecture selection process, we employed an early stopper for Scenario 0 that terminated the training when there was no improvement in the validation metrics for more than 10 epochs.

In Table 5, we collect the simulation results for scenario 0. The MSEs, RMSEs and MAEs are expressed in absolute values, the MAPEs in percentage points, and the average training time per epoch in second [s]

Table 5: Validation error metrics and computation time for one-hour-ahead prediction: Scenario 0.

Central dataset size	MSE	RMSE	MAE	MAPE	Time per epoch [s]
2	0.00013	0.01158	0.00468	29.046	1.85
5	0.00012	0.01113	0.00308	9.068	6.01
8	0.00042	0.02067	0.00611	9.734	6.19
11	0.00028	0.01681	0.00437	8.561	8.18
14	0.00022	0.01514	0.00390	7.500	10.52
17	0.00023	0.01519	0.00383	6.850	12.56
20	0.00022	0.01498	0.00387	9.017	14.59
23	0.00019	0.01388	0.00330	7.144	16.82

Table 5 highlights that the overall performance of the centralized setting is very good, and that it remains almost constant for more than five clients with no evident variation in any of the metrics. The poor results in the two-client case could be the result of substantially different consumption patterns.

5.2 Scenario A: standard federated learning setting

We designed Scenario A to compare the 'centralized setting' in Scenario 0 with a FL setting, and to obtain a reference point for the other FL scenarios. Scenario A uses the NN architecture presented in (Marino et al., 2016) and does not apply privacy-preserving techniques. Furthermore, we did not impose data correlation among the clients.

Table 6 presents the simulation results for Scenario A. The error metrics are expressed in absolute values and the average training time per epoch is expressed in seconds [s].

Table 6: Validation error metrics and computation time for one-hour-ahead prediction: Scenario A.

Federation size	MSE	RMSE	MAE	MAPE	Time per round [s]
2	0.00015	0.01240	0.00516	30.1461	3.13
5	0.00022	0.01496	0.00468	16.2269	11.54
8	0.00058	0.02407	0.00745	11.9892	10.72
11	0.00042	0.02049	0.00538	10.1082	13.39
14	0.00035	0.01872	0.00542	10.1077	18.58
17	0.00032	0.01787	0.00469	8.5392	21.05
20	0.00031	0.01775	0.00479	11.2933	25.10
23	0.00028	0.01701	0.00478	10.8257	29.39

Table 6 highlights that performance of FL models varies depending on the federation size. While MSEs, MAEs and RMSEs remain almost constant, there is a clear improvement in MAPEs. These results are in line with those by Savi et al. (Savi and Olivadese, 2021) and Fekri et al. (Fekri et al., 2021) and indicate that larger federation sizes lead to more accurate FL models.

To better illustrate this effect, we plot how the MAPEs evolved for the eight federation sizes along the training rounds in Figure 4. Overall, we can observe a *quasi-exponential* decrease over the 300 rounds, approaching final values between 6.8 and 29, which indicate reasonably good forecasts (Lewis, 1982).

In comparison to Scenario 0, we can observe an average performance decrease between 20% to 40%. FL appears to perform significantly worse than a 'centralized' setting, which is in line with other comparable studies (Briggs et al., 2021a; Husnoo et al., 2022; Lin et al., 2022).

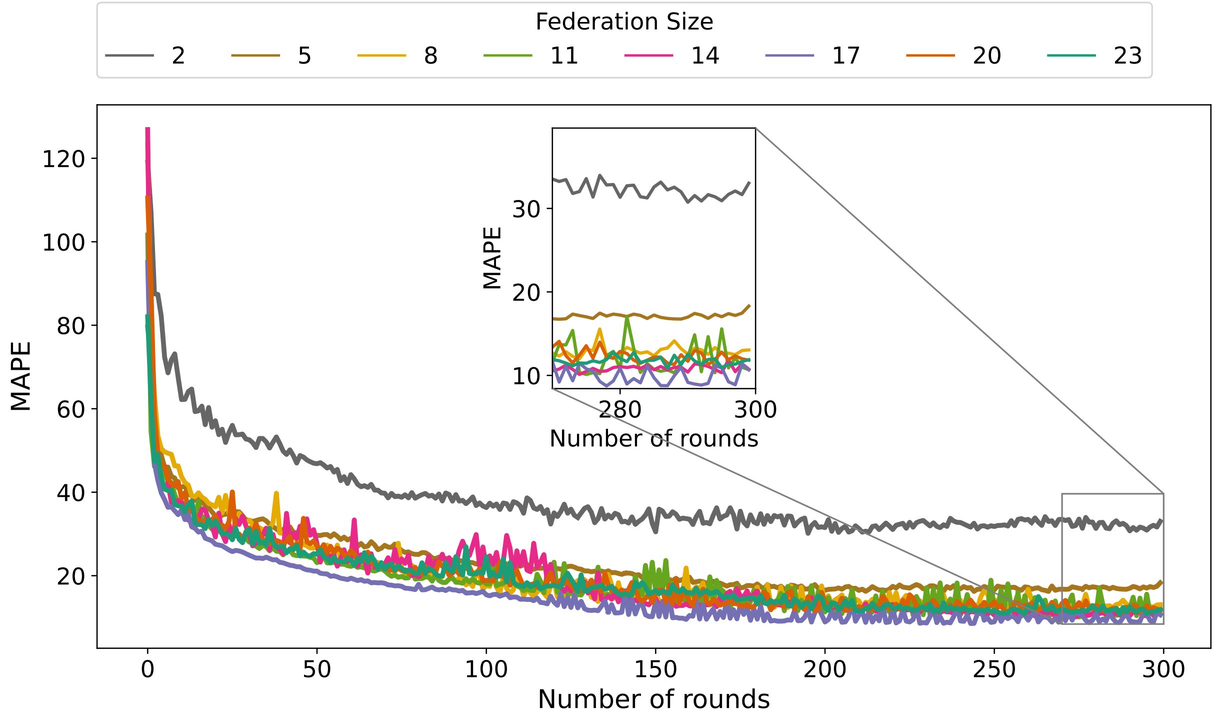


Figure 4: Validation Mean Absolute Percentage Error (MAPE) per federation size in terms of training rounds for scenario A.

Table 6 also highlights a trade-off between accuracy and computational time for federation size. As the number of clients increases, so does performance, but also computation time. This trade-off can present an important limitation for the use of FL.

5.3 Scenario B: standard federated learning setting with imposed correlation

In scenario B, we analyzed the performance of a standard FL setting with imposed correlation among the clients in the federation. We followed Lee et al. (Lee and Wu, 2020) and used Pearson correlation to identify and bundle clients (or LCLids) by correlated data. This way of bundling differs differs from the dominant k-means approach in prior literature and offers a more direct and simple view of the correlation between clients. More specifically, we pre-filtered our dataset for specific ACORNs (H and L). For these ACORNS, we then calculated all possible non-repeated combinations and calculated their correlations. For each federation size, we selected those combinations of clients with the highest correlations.

We present the simulation results for Scenario B in Table 7. The error metrics and the correlation rate are both expressed in absolute values. We omit the computation time because it was basically the same as in scenario A (5.2).

Table 7: Validation error metrics and correlation rates for one-hour-ahead prediction: scenario B.

Federation size	MSE	RMSE	MAE	MAPE	Correlation rate
2	0.00002	0.00463	0.00170	4.54	0.62
5	0.00015	0.01238	0.00373	9.77	0.51
8	0.00022	0.01513	0.00426	8.91	0.49
11	0.00021	0.01465	0.00402	8.23	0.45
14	0.00020	0.01429	0.00390	8.66	0.42
17	0.00032	0.01805	0.00465	8.22	0.37
20	0.00029	0.01726	0.00428	8.38	0.34
23	0.00026	0.01640	0.00432	9.95	0.31

FL with imposed correlation performed better in almost every metric than FL without imposed correlation (Scenario A). The MSEs decreased by an average 35.87%; RMSEs by 21.81%; MAEs by 25.57% and the MAPEs by 27.61%. They nevertheless still trail Scenario 0 by 6.35% on average. Moreover, these values are subject to some caveats. Our model with two clients had a correlation rate of 0.62, which led to a 75% better performance than the two-client case in Scenario A. Moreover, the performance of the model with 17 clients was worse than the same model in Scenario A, and 45% of the error metrics in Scenario B were better than those in scenario 0.

These results align well with similar studies, such as (Biswal et al., 2021; Fekri et al., 2021; He et al., 2021) or (Savi and Olivadese, 2021), where the application of k-means to cluster customers leads to performance improvements between 10% and 15%.

Overall, scenario B suggests that clustering based on Pearson correlation among the clients in a federation can substantially improve the performance of FL-based STLF. Specifically, utilities, energy providers, and DSOs could leverage simple socio-economic factors (ACORNS) and historical, individual smart meter data to cluster their residential customers into correlated groups. Each cluster can use a different FL model to reduce imbalance costs for inaccurate forecasts and offer tailored demand-side management programs.

5.4 Scenario C: standard federated learning setting with a more complex neural network architecture

In scenario C, we explore how a more complex NN architecture ((Khan et al., 2020)) impacts the performance of FL-based STLF. The motivation for scenario C is rooted in the trend to use ever more complex machine learning architectures in the hope of catching patterns invisible to less complex architectures. At the same time, it is unclear whether larger architectures increase performance.

To account for the size of the model in (Khan et al., 2020) and its computational burden, we implemented three modifications to the set-up of our simulation environment. The first modification concerns the GPUs. For each of the Nvidia Tesla allocated on the HPC, we created two virtual cards, resulting in four cards we could use for our simulation. The second modification is related to the batch size, which we increased from 100 to 200. Increasing the batch size can help to prevent or limit overfitting since there are more data entries available to compute the loss of the model. Finally, we modified the model in (Khan et al., 2020) by transforming the initially proposed LSTM layers to CuDNNLSTM (Appleyard et al., 2016). The transformation enabled the LSTMs to use the Compute Unified Device Architecture (CUDA) kernel of our Tesla GPUs.

The simulation results of scenario C are presented in Table 8. The results clearly indicate the increased computational costs of training a FL model with a complex architecture. The computational time is almost twice as high as in scenarios A and B. On the other hand, the performance of the model with the more complex architecture was worse that of the smaller model's for all federation sizes and all metrics, ranging from 50% up to 142%.

These results suggest a clear case of overfitting. Overfitting is generally defined as the lack of generalization of a model. An overfitted model crosses the line between learning tendencies or patterns and *memorizing* the data received as input.

Figure 5 provides a visualization of this overfitting. The performance on the training subset is represented by the solid lines, while the performance on the validation subset

Table 8: Validation error metrics and computation time for one-hour-ahead prediction: scenario C.

Federation size	MSE	RMSE	MAE	MAPE	Time per round [s]
2	0.00024	0.01550	0.00720	31.50674	6.25
5	0.00052	0.02289	0.01282	33.42653	21.10
8	0.00117	0.03433	0.01754	20.92209	20.43
11	0.00115	0.03398	0.01495	21.93438	30.34
14	0.00087	0.02955	0.01404	18.44877	34.52
17	0.00077	0.02783	0.01080	13.80498	40.59
20	0.00081	0.02858	0.01435	24.28874	50.19
23	0.00061	0.02486	0.01059	19.02717	59.76

is visualized by the dotted lines. The dotted lines begin to increase again after round 120, whereas the solid lines decrease as the model is over-fitted to the training data. .

In effect, scenario C offers a cautionary tale for utilities, energy providers, and DSOs that want to use FL for short-term load forecasting. Not only are more complex FL architectures more expensive and detrimental to the environment (Hao, 2019), they are also more sensitive to handle.

5.5 Scenario D: privacy-preserving federated learning setting with differential privacy

Scenario D focuses on adding DP to FL and how this impacts the performance of FL-based STLF. Furthermore, we compare two *flat clipping* approaches: fixed and adaptive clipping, as described in subsection 4.2.

In scenarios A and B, we used Marino et al.'s model (Marino et al., 2016) as the baseline architecture. Encoder-decoder architectures can cope well with outliers due to their capacity to abstract information into the latent space. This capacity is very beneficial for FL where different clients can have substantially different data points. However, we found that these architectures are substantially more vulnerable to noise than standard stacked LSTM networks. One reason for this vulnerability could be that they compact information from a higher dimensional space into a smaller one. Adding noise to the

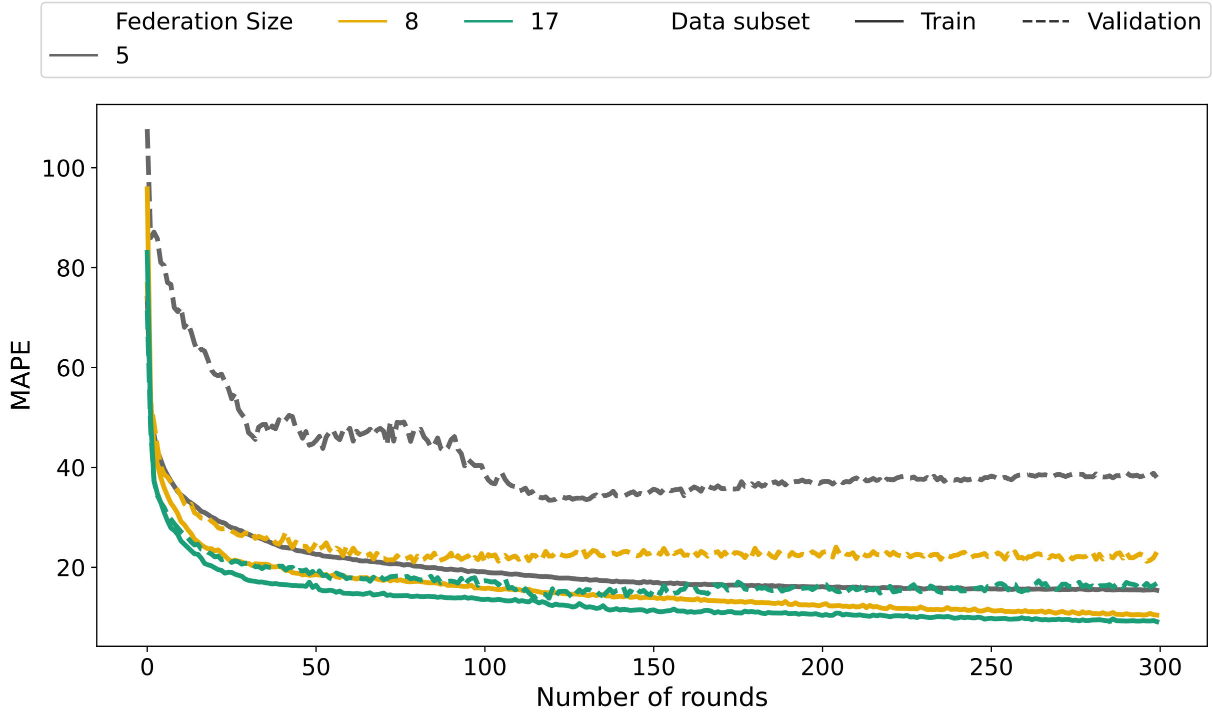


Figure 5: Validation and Training MAPEs for federation sizes 5,8, and 17 in Scenario C.

weights of this latent space will have a multiplicative effect on the model's output in the decoder phase. To avoid such encoder-decoder noise problems for our DP simulation, we changed the architecture in Scenario D to a two-layer LSTM with 50 neurons each, and a final dense layer as in He et al. (He et al., 2021).

DP offers two approaches to obtain a high privacy budget given a defined amount of noise: reduce the ratio of clients that participate in each training round (Q), retrain the model locally for several epochs on client data, find a lower δ, and/or increase the noise scale (z). For Scenario D, we employed a ratio of $Q = 0.1$. With $Q = 0.1$, a total of 100 clients and without the addition of privacy preserving techniques, our model had a MAE of 0.00300, a MSE of 0.012, a RMSE of 0.01114, and a MAPE of 8.3846, which matches results in Scenario A.

Moreover, we considered recommendations by Zhao et al. (Zhao et al., 2021) and Xu et al. (Xu et al., 2021) to introduce local re-training. Specifically, they propose to conduct several local training rounds on each client between each aggregation with DP to better fit the local models. Yet, we found that these repeated rounds didn't improve performance so we chose to use just one local training round. However, we did optimize the δ to $\delta = 4e^{-3}$ as proposed by Zhao et al. (Zhao et al., 2021).

The first strategy we implemented was fixed clipping following the two main steps in McMahan et al. (McMahan et al., 2018). In the first step, we determined the lowest possible clipping value (S) as being too low clipping values can negatively affect the convergence rate as they clip all values bigger than S. We treated S as a hyper-parameter and used an iterative approach to find the lowest possible clipping value. Specifically, we followed McMahan et al. (McMahan et al., 2018) and used iterative steps of 0.1 for S, starting with $S = 0.1$ until $S = 0.7$. We present the error metrics for the different S values in Table 9. [4].

Based on these iterations, we selected $S \approx 0.3$ as our fixed clipping value. It is the lowest clipping value with comparatively good error metrics and the marginal increase in error metrics from lowering S increases disproportionately below ≈ 0.3.

Table 9: Validation error metrics for different clipping values for one-hour-ahead prediction with the sample client ratio $Q = 0.1$ and total number of clients $w = 100$: scenario D.

S	MSE	RMSE	MAE	MAPE
0.10	0.00043	0.02094	0.00628	10.69357
0.20	0.00035	0.01884	0.00502	8.89023
0.30	0.00038	0.01969	0.00496	8.00244
0.40	0.00038	0.01963	0.00486	7.71642
0.50	0.00039	0.01978	0.00493	7.92688
0.60	0.00034	0.01869	0.00477	7.81763
0.70	0.00036	0.01915	0.00484	7.53057

Once we had identified the lowest possible clipping value S, the second step was to identify a tolerable level of noise. With $S = 0.3$, a total number of clients $w = 100$, and $Q = 0.1$, we applied $\mathbb{S} = {}^{S}/_{Qw}$ to calculate the standard deviation of the noise level $\sigma = z \cdot \mathbb{S}$. Similarly with the approach that we took with S, we treated z as a hyper-parameter and ranged it from 0.1 to 0.9

[4] Setting a fixed value for the clipping slows the training process significantly. The values in Table 9 are the validation metrics after 2000 communication rounds. Without any clipping strategy, the models converge at an earlier rate (see figure 4)

Table 10: Exploration of the different noise levels, in bold the hyper-parameter **z** that defines the amount of noise.

$\mathbf{Qw}$	$\mathbf{S}$	$\mathbb{S} = {}^{\mathbf{S}}/_{\mathbf{Qw}}$	$\mathbf{z}$	$\boldsymbol{\sigma} = \mathbf{z} \cdot \mathbb{S}$
10	0.3	0.03	0.1	0.003
10	0.3	0.03	0.2	0.006
10	0.3	0.03	0.3	0.009
10	0.3	0.03	0.4	0.012
10	0.3	0.03	0.5	0.015
10	0.3	0.03	0.6	0.018
10	0.3	0.03	0.7	0.021
10	0.3	0.03	0.8	0.024
10	0.3	0.03	0.9	0.027

In Table 10, we present the performance metrics for each of the z variations. Each of the explored z values represents a different level of noise added to the federated model. Intuitively, there is a trade-off between the amount of noise and performance, whereby more noise (increase in z) reduces performance. This trade-off dynamic is clear from the error metrics in Table 11. Nevertheless, the overall error metrics for DP based on fixed clipping are generally low and indicate good forecasting performance.

Concurrently, more noise also means better privacy, as indicated by the increasing privacy guarantees in column three of Table 11. We calculated these guarantees using the Rényi Differential Privacy Accountant (Mironov, 2017). The highest amount of noise we examined (z=0.9) provides a privacy guarantee of (4.2, $4e^{-3}$), which is close to perfect privacy ($\epsilon = 0$). In effect, scenario D demonstrates that adding DP to FL maintains comparatively good performance and offers high privacy guarantees.

The second clipping strategy that we analyzed is adaptive clipping. With adaptive clipping, clipping value are calculated automatically. To evaluate this approach, we used Andrew et al.'s adaptive clipping implementation (Andrew et al., 2021), in which the algorithm iteratively (per communication round) adjusts the norm clip, trying to approximate it to a predefined quantile (0.5 in our case).

Table 11: Validation error metrics with $S = 0.3$ and a varying noise scale **z** from 0.1 to 0.9 for one hour-ahead-prediction with the sample client ratio $Q = 0.1$ and total number of clients $w = 100$ after one epoch of local training.

Noise scale (z)	Privacy Guarantee (ϵ, δ)	MSE	RMSE	MAE	MAPE	Timer per round [s]
0.1	(911,$4e^{-3}$)	0.00010	0.00946	0.00272	7.5426	86.74
0.2	(190,$4e^{-3}$)	0.00010	0.00957	0.00312	8.8930	85.11
0.3	(69.3,$4e^{-3}$)	0.00010	0.00959	0.00309	8.4391	87.48
0.4	(32.4,$4e^{-3}$)	0.00010	0.00962	0.00321	9.1156	84.66
0.5	(17.9,$4e^{-3}$)	0.00011	0.00971	0.00340	9.7164	88.52
0.6	(11.2,$4e^{-3}$)	0.00011	0.00972	0.00344	9.9693	84.28
0.7	(7.58,$4e^{-3}$)	0.00011	0.00979	0.00354	10.0378	81.46
0.8	(5.5,$4e^{-3}$)	0.00013	0.01075	0.00519	15.6755	82.08
0.9	(4.2,$4e^{-3}$)	0.00011	0.00991	0.00372	10.6031	87.48

This data quantile approximation expends privacy budget as it queries the data. To prevent this *privacy leakage* Andrew et al. (Andrew et al., 2021) propose to add noise during the approximation. This noise (σ_b) is defined by 0.05 times the number of clients per round, in our case $\sigma_b = 0.5$. This addition of noise has a slight affect on the total privacy guarantee of the model. It results in increased effective noise as $z_\Delta = (z^{-2} - (2\sigma_b)^{-2})^{-1/2}$.

Figure 6 highlights the adaptive adjustments of the clipping value over the training rounds. There is a sharp increase in the clipping norm at the beginning of the training rounds due to the low initial clipping value $C^0 = 0.1$. Such a low quantile allows only a few data points to participate in selecting the clipping value. The smaller the quantile, the fewer data points participate and thus, it is more difficult to estimate the optimal clipping value.

As in our case, the adaptive clipping algorithm may overshoot as a result and increase the clipping norm to higher values. After this overshot, the adaptive clipping algorithm correctly approximates the optimal clipping value $S \approx 0.2$.

We present the simulation results for adaptive clipping in Table 12. On average, adaptive clipping outperformed fixed clipping by 9%. Moreover, the privacy guarantee is close to perfect privacy (3.9,$4e^{-3}$)

Adaptive clipping appears not only more attractive from a performance and privacy perspective. It is also easier to use in terms of performance and privacy. Fixed clipping

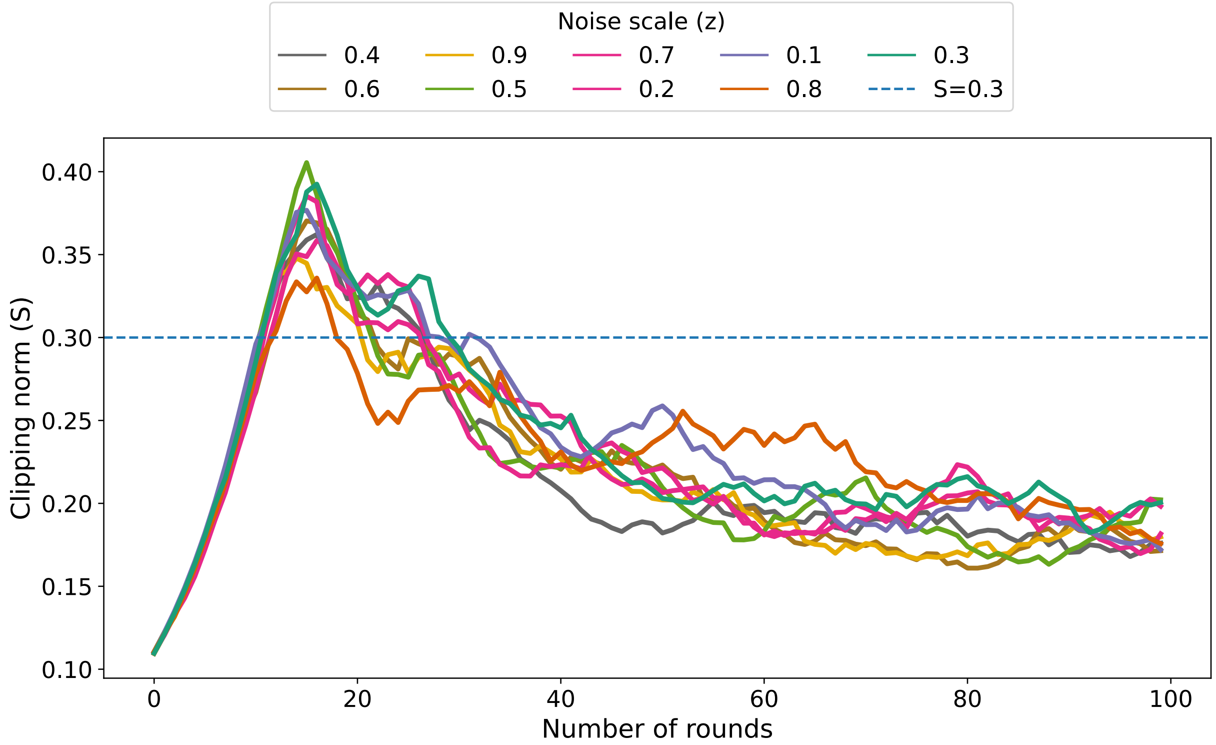


Figure 6: Evolution of the adaptive clipping norm at different noise levels z (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) using as initial clipping value $C^0 = 0.1$ and the step factor for the geometric updates $\eta C = 0.2$.

requires an initial and computationally expensive manual step to identify an appropriate clipping value, whereas, in adaptive clipping, this value is calculated automatically in the training rounds. Thus, DP with adaptive clipping presents the more convenient choice for residential STLF.

Table 12: Validation error metrics with adaptive clipping at different noise levels from 0.1 to 0.9 using as initial clipping value $C^0 = 0.1$ and the step factor for the geometric updates $\eta C = 0.2$ for one hour ahead prediction with the sample client ratio $Q = 0.1$ and total number of clients $w = 100$ after one epoch of local training.

Noise scale (z)	Effective noise (z_Δ)	Privacy Guarantee (ϵ, δ)	MSE	RMSE	MAE	MAPE	Time per round [s]
0.1	0.100	(910.0,$4e^{-3}$)	0.00010	0.00936	0.00276	7.9966	84.39
0.2	0.200	(189.4,$4e^{-3}$)	0.00010	0.00930	0.00260	7.3866	88.41
0.3	0.300	(68.7,$4e^{-3}$)	0.00009	0.00930	0.00257	7.0985	85.30
0.4	0.402	(31.9,$4e^{-3}$)	0.00010	0.00945	0.00292	8.2810	86.92
0.5	0.504	(17.5,$4e^{-3}$)	0.00010	0.00948	0.00301	9.0461	88.57
0.6	0.607	(10.8,$4e^{-3}$)	0.00010	0.00955	0.00302	8.8343	86.27
0.7	0.711	(7.2,$4e^{-3}$)	0.00010	0.00961	0.00317	9.4312	87.68
0.8	0.817	(5.2,$4e^{-3}$)	0.00010	0.00955	0.00325	9.6126	88.27
0.9	0.924	(3.9,$4e^{-3}$)	0.00010	0.00955	0.00319	9.2953	87.93

The results we present in Tables 11 and 12 are those after the local training round suggested by Zhao et al. (Zhao et al., 2021). Unlike Zhao et al. (Zhao et al., 2021), who worked with five local training round, we used only one as additional rounds did not significantly improve performance (Figure 7). Nevertheless, clients profited from local training with negligible computational overhead.

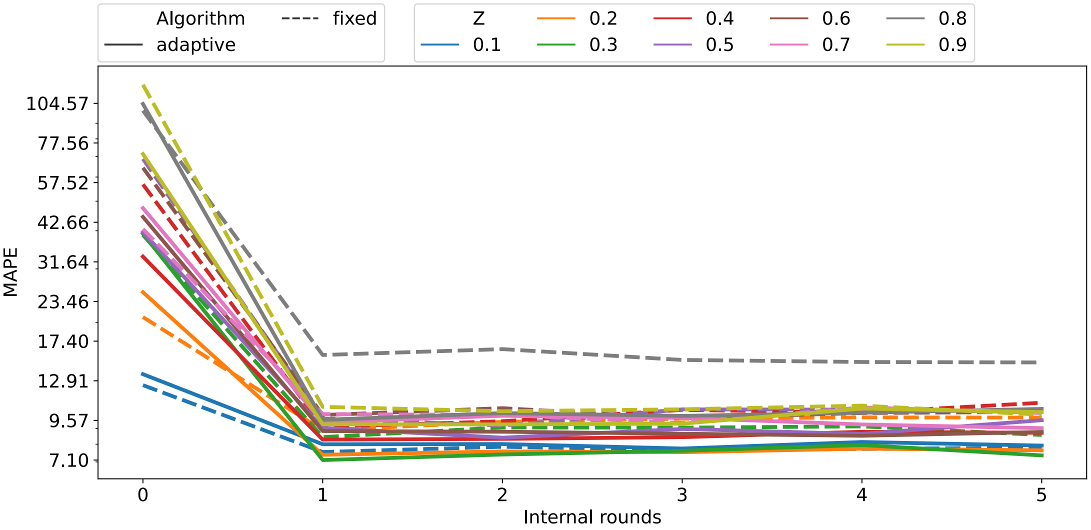


Figure 7: Validation Mean Absolute Percentage Error (MAPE) per local training epoch for adaptive and fixed DP.

5.6 Scenario E: privacy-preserving federated learning setting with secure aggregation

In this scenario, we examine SecAgg as an alternative technique to add privacy to FL. Whereas DP adds random noise to model updates, SecAgg targets the communication and aggregation of the clients' model updates. Hence, there is no trade-off as in scenario D, where it is important to find an adequate noise level.

Similar to scenarios A, B and C, we present the simulation results for the eight federation sizes in Table 13. We express the error metrics in absolute values and the average computation time in seconds [s]. Furthermore, we complement the results with Figure 8. It depicts the MAPE, following a similar curve as in Scenario A.

Table 13 shows that the use of SecAgg affects computation time only marginally. As SecAgg does not add any noise, it also provides less burden than DP. Consequently, SecAgg presents a more performant alternative for residential STLF with the cost of an extra 30% of computation time. However, it is important to note that SecAgg does not provide complete privacy because latent patterns could still point toward the original data subject. More specifically, Model Inversion (MI) attacks could reconstruct the original training data from the model parameters (Fredrikson et al., 2015).

Table 13: Error metrics and computation time for one-hour-ahead prediction using SecAgg: scenario E on test set.

Federation size	MSE	RMSE	MAE	MAPE	Time per round [s]
2	0.00017	0.01324	0.00532	31.01177	4.54
5	0.00018	0.01348	0.00431	15.60893	13.23
8	0.00060	0.02457	0.00759	12.28532	13.34
11	0.00039	0.01996	0.00523	9.65965	18.21
14	0.00034	0.01864	0.00503	9.67057	22.25
17	0.00033	0.01820	0.00466	8.25973	26.70
20	0.00033	0.01836	0.00522	12.88359	34.64
23	0.00028	0.01683	0.00453	10.19247	38.10

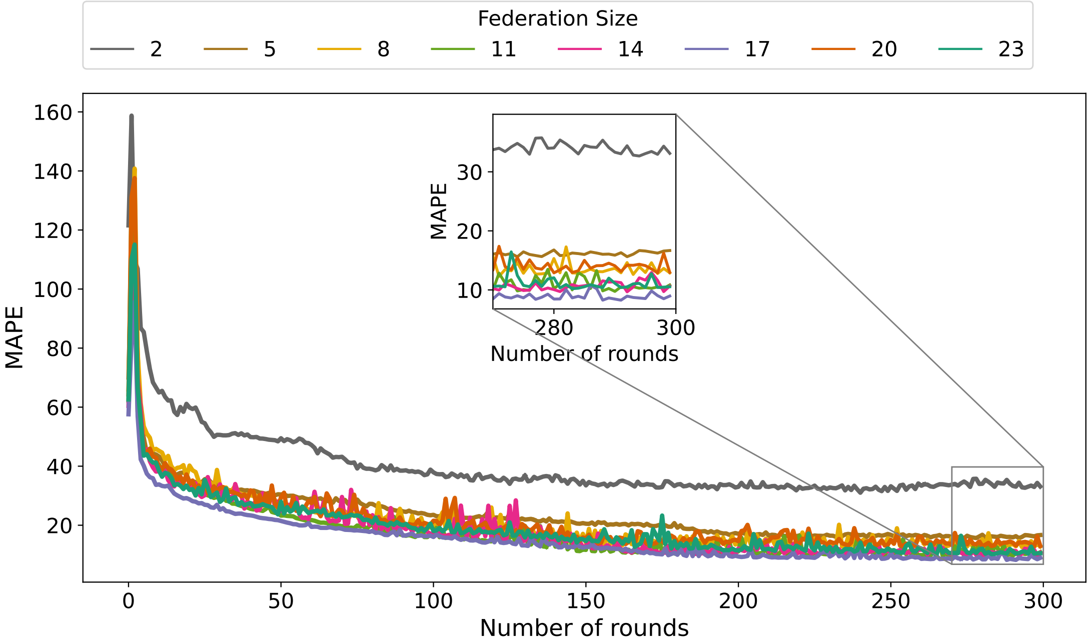


Figure 8: Validation Mean Absolute Percentage Error (MAPE) per LCLids federation size in terms of training rounds for Scenario E.

5.7 Comparison across the scenarios

We summarize our results for scenarios 0, A, B, C, and E in Figures 9 and 10. We omitted scenario D from these figures because in scenario D we only varied the noise scale and not the federation size.

Overall, the two figures suggest an inherent trade-off between performance and privacy in residential STLF. Yet, FL models can successfully mediate this trade-off and provide high levels of performance and privacy, especially when trained on correlated data, avoid unduly complex architectures, and employ SecAgg.

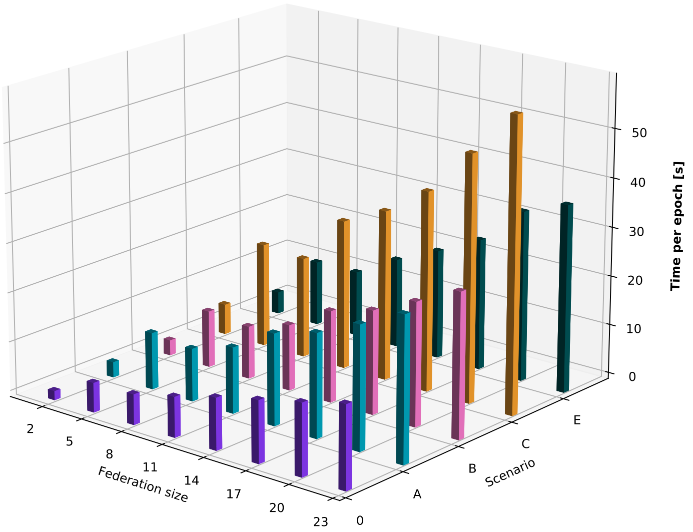
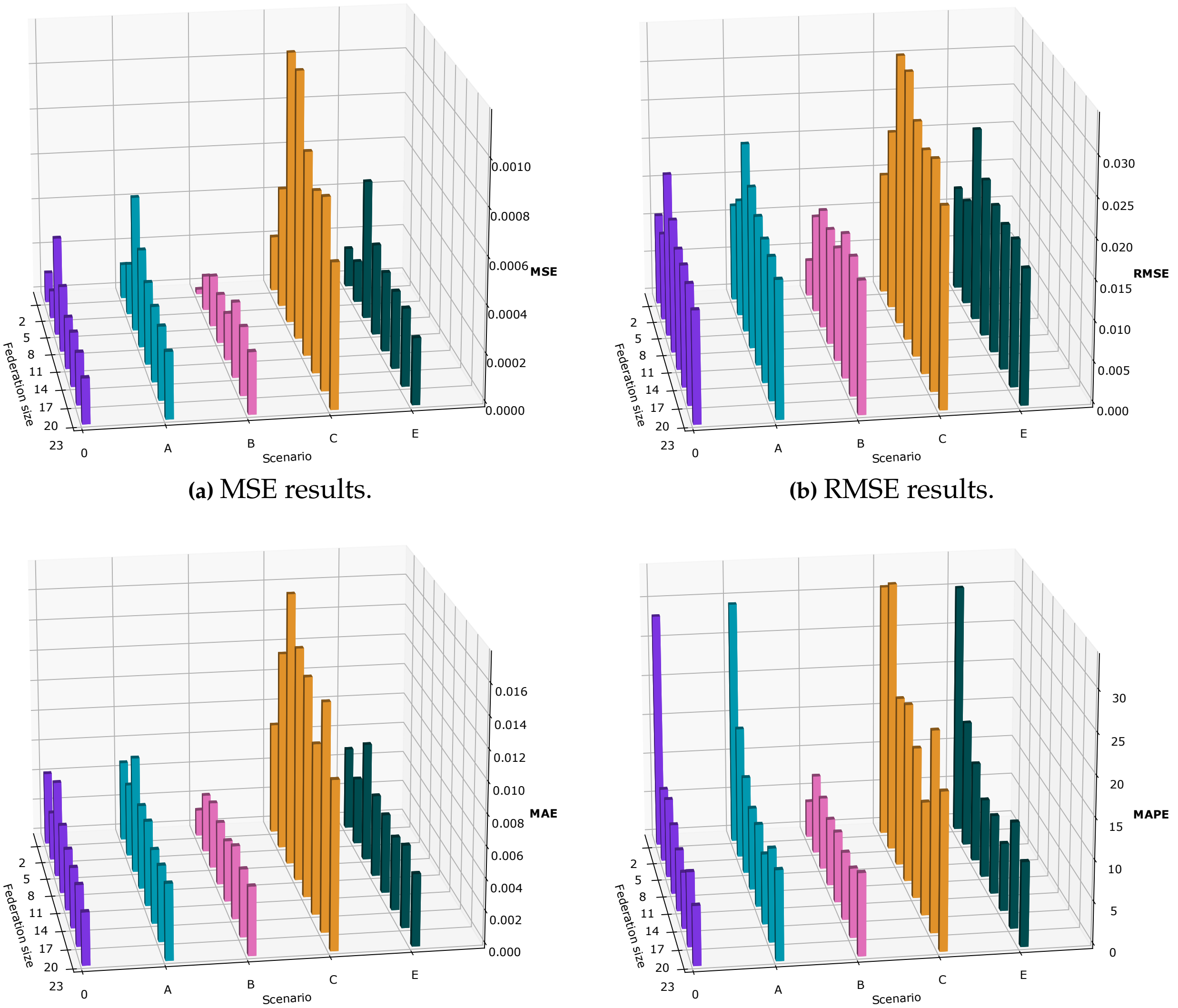


Figure 9: Comparison of computation time across Scenarios 0, A, B, C, and E.

6 Conclusions

This paper analyses the use of FL and its combination with privacy preserving techniques for short-term forecasting of individual residential loads. Such a combination offers an innovative approach to accommodate both accuracy and privacy. In particular, it allows those who depend on accurate forecasts of residential loads (such as utilities, energy providers, and DSOs) to train in a collaborative fashion forecasting models with granular smart meter data without having to share this data.

Our analysis builds on historical smart meter data and consists of six scenarios. While the first two scenarios set the baseline scenarios, each of the subsequent four scenarios have a particular analytical focus. Specifically, these scenarios investigate the effects of data correlation, neural network architecture complexity, differential privacy, and secure aggregation on performance, computation time, and privacy guarantee levels. In each scenario, we also explore the effects of different federation sizes. From our analysis, we can posit the following:



(a) MSE results.

(b) RMSE results.

(c) MAE results.

(d) MAPE results.

Figure 10: Comparison of evaluation metrics across Scenarios 0, A, B, C, and E.

1. Collaborative training of AI models with federated learning reduces forecasting accuracy as compared to a ’centralized’ setting. However, it makes it easier to account for data privacy concerns through the addition of privacy-preserving techniques.

2. As the number of participating clients (smart meters) in a federation increases, forecasting accuracy tends to also increase. However, while a greater number of clients leads to greater accuracy, this also implies higher computational costs that may no always be justified.

3. Customer segmentation with Pearson correlation along socio-economic factors (e.g., with the ACORN methodology) substantially improves forecasting accuracy for FL models.

4. Complex neural network architectures imply high computational costs, difficulties in handling the architecture, and a potential risk of overfitting. It is thus important to balance accuracy and usability when selecting of model architectures.

5. Complementing federated learning with differential privacy or secure aggregation does not significantly reduce forecasting accuracy but does enable very high levels of privacy.

6. Adaptive and fixed clipping approaches to differential privacy provides similar performance. Adaptive clipping is easier to use as it does not require manual preselection of good clipping values, and it facilitates faster model convergence.

7. Combining autoencoder architectures with DP complicates the training of FL models. The design of these architectures magnifies the noise added by DP, which restricts the training process.

8. Secure aggregation is superior to DP in terms of usability, performance and computational burden. It can be added as a simple plug-and-play component, does not reduce performance by adding noise, and permits faster training.

Overall, our analysis suggests that a combination of federated learning with privacy-preserving techniques can be a highly promising alternative for residential short-term load forecasting. However, is not free from technical challenges. Differential privacy requires careful configuration of noise size, clipping values and client ratios to balance accuracy and privacy. Secure aggregation does not require such configuration but its cryptographic set-up can also be challenging as well. Furthermore, computational costs limit the number of clients that can be used for training.

More broadly, our study contributes to a better understanding of the use of FL and privacy-preserving techniques for residential short-term load forecasting. It makes an important contribution to the growing literature on the applications of federated learning in electric power systems by testing different NN under distributed settings, examining the implications of privacy preserving techniques, and identifying technical challenges in using FL.

Naturally, our analysis is not free from limitations. In particular, computational costs have considerably limited the size of our federations. Even though larger federation sizes may result in somewhat different results, nevertheless we believe that our overall results are robust, as we have explored several settings in terms of: number of clients, baseline NN architectures, and dataset characteristics.

Further research may nevertheless want to (1) assess larger federation size settings with additional correlation indicators, such as the existence of distributed energy resources (i.e., photovoltaics, electric vehicles, or home energy management systems), (2) investigate data input disruptions produced by hostile agents or errors caused by malfunctions of a smart metering device, and (3) examine other, innovative NN architectures with attention mechanisms and multi-variate input data. After all, FL is highly collaborative and iterative and perfect data and operation may not always be possible in real-world applications.

Credit authorship contribution statement

Conceptualization, J.D.F, S.P.M, C.L; Methodology, J.D.F, S.P.M, C.L; Data Curation, J.D.F, S.P.M; Writing - Original Draft, J.D.F, S.P.M, C.L; Software J.D.F; Supervision A.R, G.F.; Writing - Review & Editing, A.R, G.F.; Visualization, J.D.F, S.P.M.; Funding acquisition, G.F. All authors have read and agreed to the published version of the manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

We would like to acknowledge Tom Josua Barberau and Orestis Papageorgiou for their valuable feedback on the first draft of this paper.

Moreover, we would like to acknowledge the support of the European Union (EU) within its Horizon 2020 programme, project MDOT (Medical Device Obligations Taskforce), Grant agreement 814654. Additionally, this project was supported by the Kopernikus-project “SynErgie” of the German Federal Ministry of Education and Research (BMBF), by PayPal and the Luxembourg National Research Fund FNR (P17/IS/13342933/PayPal-FNR/Chair in DFS/Gilbert Fridgen) as well as by the Luxembourg National Research Fund (FNR) – FiReSpARX Project, ref. 14783405.

References

Abadi, M., A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang (2016). “Deep learning with differential privacy”. In: *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318.

Alfares, H. K. and M. Nazeeruddin (2002). “Electric load forecasting: Literature survey and classification of methods”. In: *International Journal of Systems Science* 33.1, pp. 23–34. DOI: 10.1080/00207720110067421. eprint: https://doi.org/10.1080/00207720110067421. URL: https://doi.org/10.1080/00207720110067421.

Andrew, G., O. Thakkar, B. McMahan, and S. Ramaswamy (2021). “Differentially private learning with adaptive clipping”. In: *Advances in Neural Information Processing Systems* 34.

Appleyard, J., T. Kocisky, and P. Blunsom (2016). *Optimizing Performance of Recurrent Neural Networks on GPUs*. arXiv: 1604.01946 [cs.LG].

Ardabili, S., A. Mosavi, and A. R. Várkonyi-Kóczy (2019). “Advances in machine learning modeling reviewing hybrid and ensemble methods”. In: *International Conference on Global Research and Education*. Springer, pp. 215–227.

Atrias (2021). *The central hub in providing information in the energy market.* Accessed: 2021-05-28. URL: https://www.atrias.be/.

Barbosa, P., A. Brito, and H. Almeida (2016). “A Technique to provide differential privacy for appliance usage in smart metering”. In: *Information Sciences* 370-371, pp. 355–367. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2016.08.011. URL: https://www.sciencedirect.com/science/article/pii/S0020025516305862.

Bennett, C. J. (2018). *Regulating Privacy*. Cornell University Press. DOI: doi:10.7591/9781501722134. URL: https://doi.org/10.7591/9781501722134.

Bielecki, S., T. Skoczkowski, L. Sobczak, J. Buchoski, L. Maciąg, and P. Dukat (2021). "Impact of the Lockdown during the COVID-19 Pandemic on Electricity Use by Residential Users". In: *Energies* 14.4. ISSN: 1996-1073. DOI: 10.3390/en14040980. URL: https://www.mdpi.com/1996-1073/14/4/980.

Biswal, M. K., A. S. M. Tayeen, and S. Misra (2021). "AMI-FML: A Privacy-Preserving Federated Machine Learning Framework for AMI". In: *arXiv* abs/2109.05666.

Bonawitz, K., V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth (2017). "Practical Secure Aggregation for Privacy-Preserving Machine Learning". In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. CCS '17. Dallas, Texas, USA: Association for Computing Machinery, 1175–1191. ISBN: 9781450349468. DOI: 10.1145/3133956.3133982. URL: https://doi.org/10.1145/3133956.3133982.

Briggs, C., Z. Fan, and P. Andras (2021a). *Federated Learning for Short-term Residential Energy Demand Forecasting*. arXiv: 2105.13325 [cs.LG].

Briggs, C., Z. Fan, and P. András (2021b). "Federated Learning for Short-term Residential Energy Demand Forecasting". In: *arXiv* abs/2105.13325.

CACI (2014). *The acorn user guide*. Tech. rep. CACI.

Chhachhi, S. and F. Teng (2021). "Market Value of Differentially-Private Smart Meter Data". In: *2021 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5. DOI: 10.1109/ISGT49243.2021.9372228.

Cho, K., B. van Merriënboer, D. Bahdanau, and Y. Bengio (2014). "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches". In: *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Doha, Qatar: Association for Computational Linguistics, pp. 103–111. DOI: 10.3115/v1/W14-4012. URL: https://aclanthology.org/W14-4012.

Chollet, F. et al. (2015). *Keras*. https://keras.io.

Commission for Regulation of Utilities (2017). *Energy Supply Costs Information Paper*. Tech. rep. CRU17291. Accessed on 2022-02-16. Commission for Regulation of Utilities - CRU.

D., J.-M. (2019). *Smart meter data from London area*. URL: https://www.kaggle.com/jeanmidev/smart-meters-in-london.

Dwork, C. (2006). “Differential Privacy”. In: *Automata, Languages and Programming*. Ed. by M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–12.

Dwork, C. and A. Roth (2014). “The Algorithmic Foundations of Differential Privacy”. In: *Found. Trends Theor. Comput. Sci.* 9.3–4, 211–407. ISSN: 1551-305X. DOI: 10.1561/0400000042. URL: https://doi.org/10.1561/0400000042.

Eibl, G. and D. Engel (2017a). “Differential privacy for real smart metering data”. In: *Computer Science-Research and Development* 32.1-2, pp. 173–182.

Eibl, G. and D. Engel (2017b). “Differential privacy for real smart metering data”. In: *Computer Science - Research and Development* 32. DOI: 10.1007/s00450-016-0310-y.

Elhub (2021). *Netural data hub fir metering data and market processes.* Accessed: 2021-05-28. URL: https://elhub.no/en/#.

ENTSO-E (2021). *Enhanced Load Forecasting*. URL: https://www.entsoe.eu/Technopedia/techsheets/enhanced-load-forecasting.

European Commission, Directorate-General for Energy, G Küpper, M Cavarretta, A Ehrenmann, E Naffah, J Szilagyi, D Guldentops, N Rozai, and L Charlier ((2020)). *Format and procedures for electricity (and gas) data access and exchange in Member States*. Publications Office. DOI: doi/10.2833/719689.

Fekri, M. N., K. Grolinger, and S. Mir (2021). “Distributed load forecasting using smart meter data: Federated learning with Recurrent Neural Networks”. In: *International Journal of Electrical Power & Energy Systems*, p. 107669. ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2021.107669. URL: https://www.sciencedirect.com/science/article/pii/S0142061521008991.

Fredrikson, M., S. Jha, and T. Ristenpart (2015). “Model inversion attacks that exploit confidence information and basic countermeasures”. In: *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pp. 1322–1333.

Geiping, J., H. Bauermeister, H. Dröge, and M. Moeller (2020). “Inverting gradients-how easy is it to break privacy in federated learning?” In: *Advances in Neural Information Processing Systems* 33, pp. 16937–16947.

Glorot, X. and Y. Bengio (2010). “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, pp. 249–256.

Haney, A., T. Jamasb, and M. Pollitt (2009). "Smart Metering and Electricity Demand: Technology, Economics and International Experience". In: *Faculty of Economics, University of Cambridge, Cambridge Working Papers in Economics*.

Hao, K. (2019). "Training a single AI model can emit as much carbon as five cars in their lifetimes". In: *MIT Technology Review*.

He, Y., F. Luo, G. Ranzi, and W. Kong (2021). "Short-Term Residential Load Forecasting Based on Federated Learning and Load Clustering". In: *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 77–82.

Hinterstocker, M., P. Schott, and S. von Roon (2017). *Disaggregation of household load profiles*.

Hippert, H., C. Pedreira, and R. Souza (2001). "Neural networks for short-term load forecasting: a review and evaluation". In: *IEEE Transactions on Power Systems* 16.1, pp. 44–55. DOI: 10.1109/59.910780.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Husnoo, M. A., A. Anwar, N. Hosseinzadeh, S. N. Islam, A. N. Mahmood, and R. R. M. Doss (2022). "FedREP: Towards Horizontal Federated Load Forecasting for Retail Energy Providers". In: *arXiv* abs/2203.00219.

International Energy Agency (2021). *Electricity final consumption by sector, Wolrd*. Accessed on 2021-06-21. URL: https://www.iea.org/fuels-and-technologies/electricity.

Jayaraman, B. and D. Evans (2019). "Evaluating differentially private machine learning in practice". In: *28th {USENIX} Security Symposium ({USENIX} Security 19)*, pp. 1895–1912.

Kairouz, P., H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al. (2021). "Advances and open problems in federated learning". In: *Foundations and Trends® in Machine Learning* 14.1–2, pp. 1–210.

Kalimoldayev, M., A. Drozdenko, I. Koplyk, T. Marinich, A. Abdildayeva, and T. Zhukabayeva (2020). "Analysis of modern approaches for the prediction of electric energy consumption". In: *Open Engineering* 10.1, pp. 350–361. DOI: doi:10.1515/eng-2020-0028. URL: https://doi.org/10.1515/eng-2020-0028.

Kaur, H. and S. Ahuja (2017). “Time series analysis and prediction of electricity consumption of health care institution using ARIMA model”. In: *Proceedings of Sixth International Conference on Soft Computing for Problem Solving*. Springer, pp. 347–358.

Keramidas, K., A. Diaz Vazquez, M. Weitzel, T. Vandyck, M. Tamba, S. Tchung-Ming, A. Soria-Ramirez, J. Krause, R. Van Dingenen, Q Chai, et al. (2020). “Global Energy and Climate Outlook 2019: Electrification for the low carbon transition”. In: *Publications Office of the European Union, Joint Research Center: Luxembourg*.

Khalil, M., M. Esseghir, and L. Merghem (2021). “Federated Learning for Energy-Efficient Thermal Comfort Control Service in Smart Buildings”. In: *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 01–06.

Khan, Z. A., T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik (2020). “Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework”. In: *Sensors* 20.5. ISSN: 1424-8220. DOI: 10.3390/s20051399. URL: https://www.mdpi.com/1424-8220/20/5/1399.

Kim, J.-Y. and S.-B. Cho (2019a). “Electric Energy Consumption Prediction by Deep Learning with State Explainable Autoencoder”. In: *Energies* 12.4. ISSN: 1996-1073. DOI: 10.3390/en12040739. URL: https://www.mdpi.com/1996-1073/12/4/739.

Kim, T.-Y. and S.-B. Cho (2019b). “Predicting residential energy consumption using CNN-LSTM neural networks”. In: *Energy* 182, pp. 72–81.

Konečný, J., H. B. McMahan, D. Ramage, and P. Richtárik (2016). *Federated Optimization: Distributed Machine Learning for On-Device Intelligence*. arXiv: 1610.02527 [cs.LG].

Kong, W., Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang (2019). “Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network”. In: *IEEE Transactions on Smart Grid* 10.1, pp. 841–851. DOI: 10.1109/TSG.2017.2753802.

Kowarik, A., P. Stolze, O. Grondal, M. Ilves, T. Kirt, I. Jansson, and D. Wu (2016). *Report on data access and data handling*. Tech. rep. ESSnet Big Data. URL: https://ec.europa.eu/eurostat/cros/content/WP3_Report_1_1_en.

Le, T., M. T. Vo, B. Vo, E. Hwang, S. Rho, and S. W. Baik (app9204237). “Improving Electric Energy Consumption Prediction Using CNN and Bi-LSTM”. In: *Applied Sciences* 9.20. ISSN: 2076-3417. DOI: 10.3390/app9204237. URL: https://www.mdpi.com/2076-3417/9/20/4237.

Lee, C.-Y. and C.-E. Wu (2020). “Short-Term Electricity Price Forecasting Based on Similar Day-Based Neural Network”. In: *Energies* 13.17. ISSN: 1996-1073. DOI: 10.3390/en13174408. URL: https://www.mdpi.com/1996-1073/13/17/4408.

Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann.

Li, C. (2020). “Designing a short-term load forecasting model in the urban smart grid system”. In: *Applied Energy* 266, p. 114850. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2020.114850. URL: https://www.sciencedirect.com/science/article/pii/S0306261920303627.

Li, C., Z. Ding, D. Zhao, J. Yi, and G. Zhang (2017). “Building Energy Consumption Prediction: An Extreme Deep Learning Approach”. In: *Energies* 10.10. ISSN: 1996-1073. DOI: 10.3390/en10101525. URL: https://www.mdpi.com/1996-1073/10/10/1525.

Li, J., Y. Ren, S. Fang, K. Li, and M. Sun (2020a). “Federated Learning-Based Ultra-Short term load forecasting in power Internet of things”. In: *2020 IEEE International Conference on Energy Internet (ICEI)*, pp. 63–68.

Li, Q., Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He (2021). “A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection”. In: *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1. DOI: 10.1109/TKDE.2021.3124599.

Li, T., A. K. Sahu, A. Talwalkar, and V. Smith (2020b). “Federated Learning: Challenges, Methods, and Future Directions”. In: *IEEE Signal Processing Magazine* 37.3, 50–60. ISSN: 1558-0792. DOI: 10.1109/msp.2020.2975749. URL: http://dx.doi.org/10.1109/MSP.2020.2975749.

Lin, J., J. Ma, and J. Zhu (2022). “Privacy-Preserving Household Characteristic Identification With Federated Learning Method”. In: *IEEE Transactions on Smart Grid* 13, pp. 1088–1099.

Lu, Y., X. Huang, Y. Dai, S. Maharjan, and Y. Zhang (2019). “Blockchain and federated learning for privacy-preserved data sharing in industrial IoT”. In: *IEEE Transactions on Industrial Informatics* 16.6, pp. 4177–4186.

Lusis, P., K. R. Khalilpour, L. Andrew, and A. Liebman (2017). “Short-term residential load forecasting: Impact of calendar effects and forecast granularity”. In: *Applied Energy* 205, pp. 654–669. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.201

7.07.114. URL: https://www.sciencedirect.com/science/article/pii/S0306261917309881.

Maltais, L.-G. and L. Gosselin (2021). "Forecasting of short-term lighting and plug load electricity consumption in single residential units: Development and assessment of data-driven models for different horizons". In: *Applied Energy*, p. 118229. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2021.118229. URL: https://www.sciencedirect.com/science/article/pii/S030626192101494X.

Marino, D. L., K. Amarasinghe, and M. Manic (2016). "Building energy load forecasting using deep neural networks". In: *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*. IEEE, pp. 7046–7051.

McKenna, E., I. Richardson, and M. Thomson (2012). "Smart meter data: Balancing consumer privacy concerns with legitimate applications". In: *Energy Policy* 41, pp. 807–814. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2011.11.049. URL: https://www.sciencedirect.com/science/article/pii/S0301421511009438.

McMahan, B., E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas (2017). "Communication-efficient learning of deep networks from decentralized data". In: *Artificial intelligence and statistics*. PMLR, pp. 1273–1282.

McMahan, H. B., E. Moore, D. Ramage, and B. A. y Arcas (2016). "Federated Learning of Deep Networks using Model Averaging". In: *CoRR* abs/1602.05629. arXiv: 1602.05629. URL: http://arxiv.org/abs/1602.05629.

McMahan, H. B., D. Ramage, K. Talwar, and L. Zhang (2018). *Learning Differentially Private Recurrent Language Models*. arXiv: 1710.06963 `[cs.LG]`.

Mironov, I. (2017). "Rényi Differential Privacy". In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. DOI: 10.1109/csf.2017.11. URL: http://dx.doi.org/10.1109/CSF.2017.11.

Muñoz, A., E. F. Sánchez-Úbeda, A. Cruz, and J. Marín ((2010)). "Short-term Forecasting in Power Systems: A Guided Tour". In: *Handbook of Power Systems II*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 129–160. ISBN: 978-3-642-12686-4. DOI: 10.1007/978-3-642-12686-4_5. URL: https://doi.org/10.1007/978-3-642-12686-4_5.

Negnevitsky, M., P. Mandal, and A. K. Srivastava (2009). "An overview of forecasting problems and techniques in power systems". In: *2009 IEEE Power Energy Society General Meeting*, pp. 1–4. DOI: 10.1109/PES.2009.5275480.

Nti, I. K., M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya (2020). "Electricity load forecasting: a systematic review". In: *Journal of Electrical Systems and Information Technology* 7.1, p. 13. DOI: 10.1186/s43067-020-00021-8. URL: https://doi.org/10.1186/s43067-020-00021-8.

Petropoulos, F., D. Apiletti, V. Assimakopoulos, M. Z. Babai, D. K. Barrow, S. B. Taieb, C. Bergmeir, R. J. Bessa, J. Bijak, J. E. Boylan, et al. (2022). "Forecasting: theory and practice". In: *International Journal of Forecasting* 38.3, pp. 705–871. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2021.11.001. URL: https://www.sciencedirect.com/science/article/pii/S0169207021001758.

Pressmair, G., E. Kapassa, D. Casado-Mansilla, C. E. Borges, and M. Themistocleous (2021). "Overcoming barriers for the adoption of Local Energy and Flexibility Markets: A user-centric and hybrid model". In: *Journal of Cleaner Production* 317, p. 128323. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2021.128323. URL: https://www.sciencedirect.com/science/article/pii/S095965262102535X.

Rashed Mohassel, R., A. Fung, F. Mohammadi, and K. Raahemifar (2014). "A survey on Advanced Metering Infrastructure". In: *International Journal of Electrical Power & Energy Systems* 63, pp. 473–484. ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2014.06.025. URL: https://www.sciencedirect.com/science/article/pii/S0142061514003743.

Savi, M. and F. Olivadese (2021). "Short-Term Energy Consumption Forecasting at the Edge: A Federated Learning Approach". In: *IEEE Access* 9, pp. 95949–95969. DOI: 10.1109/ACCESS.2021.3094089.

Sehovac, L. and K. Grolinger (2020). "Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks With Attention". In: *IEEE Access* 8, pp. 36411–36426. DOI: 10.1109/ACCESS.2020.2975738.

Shamir, A. (1979). "How to share a secret". In: *Communications of the ACM* 22.11, pp. 612–613.

Shi, H., M. Xu, and R. Li (2018). "Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN". In: *IEEE Transactions on Smart Grid* 9.5, pp. 5271–5280. DOI: 10.1109/TSG.2017.2686012.

Shi, Y. and X. Xu (2022). "Deep Federated Adaptation: An Adaptative Residential Load Forecasting Approach with Federated Learning". In: *Sensors (Basel, Switzerland)* 22.

Specht, J. M. and R. Madlener (2019). “Energy Supplier 2.0: A conceptual business model for energy suppliers aggregating flexible distributed assets and policy issues raised”. In: *Energy Policy* 135, p. 110911. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2019.110911. URL: https://www.sciencedirect.com/science/article/pii/S0301421519304896.

Sutskever, I., O. Vinyals, and Q. V. Le (2014). “Sequence to sequence learning with neural networks”. In: *Advances in neural information processing systems* 27.

Taïk, A. and S. Cherkaoui (2020). “Electrical load forecasting using edge computing and federated learning”. In: *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, pp. 1–6.

Varrette, S., P. Bouvry, H. Cartiaux, and F. Georgatos (2014). “Management of an Academic HPC Cluster: The UL Experience”. In: *Proc. of the 2014 Intl. Conf. on High Performance Computing & Simulation (HPCS 2014)*. Bologna, Italy: IEEE, pp. 959–967.

Vos, M., C. Bender-Saebelkampf, and S. Albayrak (2018). “Residential Short-Term Load Forecasting Using Convolutional Neural Networks”. In: *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm 2018*. DOI: 10.1109/SmartGridComm.2018.8587494.

Wang, Y., Q. Chen, T. Hong, and C. Kang (2019). “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges”. In: *IEEE Transactions on Smart Grid* 10.3, pp. 3125–3148. DOI: 10.1109/TSG.2018.2818167.

Wood, A., M. Altman, A. Bembenek, M. Bun, M. Gaboardi, J. Honaker, K. Nissim, D. R. OBrien, T. Steinke, and S. Vadhan (2018). “Differential privacy: A primer for a nontechnical audience”. In: *Vanderbilt Journal of Entertainment & Technology Law* 21.1, pp. 209–275. URL: http://www.jetlaw.org/journal-archives/volume-21/volume-21-issue-1/differential-privacy-a-primer-for-a-non-technical-audience/.

Xian, X., X. Wang, J. Ding, and R. Ghanadan (2020). “Assisted learning: A framework for multi-organization learning”. In: *Advances in Neural Information Processing Systems* 33, pp. 14580–14591.

Xu, Y., C. Jiang, Z. Zheng, B. Yang, and N. Zhu (2021). “LSTM Short-term Residential Load Forecasting Based on Federated Learning”. In: *2021 International Conference on Mechanical, Aerospace and Automotive Engineering*.

Yan, K., X. Wang, Y. Du, N. Jin, H. Huang, and H. Zhou (2018). “Multi-Step Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy”. In: *Ener-*

gies 11.11. ISSN: 1996-1073. DOI: 10.3390/en11113089. URL: https://www.mdpi.com/1996-1073/11/11/3089.

Yang, Q., Y. Liu, T. Chen, and Y. Tong (2019a). "Federated machine learning: Concept and applications". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.2, pp. 1–19.

Yang, Q., Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu (2019b). "Federated Learning". In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 13.3, pp. 1–207. DOI: 10.2200/S00960ED2V01Y201910AIM043. eprint: https://doi.org/10.2200/S00960ED2V01Y201910AIM043. URL: https://doi.org/10.2200/S00960ED2V01Y201910AIM043.

Zhao, J., T. Jung, Y. Wang, and X. Li (2014). "Achieving differential privacy of data disclosure in the smart grid". In: *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 504–512. DOI: 10.1109/INFOCOM.2014.6847974.

Zhao, Y.-M., W. Xiao, L. Shuai, J. Luo, S. Yao, and M. Zhang (2021). "A Differential Privacy-enhanced Federated Learning Method for Short-Term Household Load Forecasting in Smart Grid". In: *2021 7th International Conference on Computer and Communications (ICCC)*, pp. 1399–1404.

Zhu, L., Z. Liu, and S. Han (2019). "Deep Leakage from Gradients". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2019/file/60a6c4002cc7b29142def8871531281a-Paper.pdf.

Research Paper 2 – *Towards a peer-to-peer residential short-term load forecasting with federated learning*

Full Title:

Towards a peer-to-peer residential short-term load forecasting with federated learning

Authors:

Joaquín Delgado Fernández, Sergio Potenciano Menci, Ivan Pavić

Published in: The 15th IEEE PES PowerTech Conference (2023)

Keywords:

Federated Learning, Peer-to-Peer, Clustering, K-means, Agent-Based Modelling, STLF

DOI:

https://doi.org/10.1109/PowerTech55446.2023.10202782

Copyright reuse license:

Abstract

The inclusion of intermittent and renewable energy sources has increased the importance of demand forecasting in the power systems. Smart meters play a critical role in modern load forecasting due to the high granularity of the measurement data. federated learning (FL) can enable accurate residential load forecasting in a distributed manner. In this regard, to compensate for the variability of households, clustering them in groups with similar patterns can lead to more accurate forecasts. Usually, clustering requires a central server that has access to the entire dataset, which collides with the decentralized nature of federated learning. In order to complement federated learning, this study proposes a decentralized peer-to-peer (P2P) strategy that employs agent-based modeling. We evaluate it in comparison to a typical centralized k-means clustering. To create clusters, we compare Euclidian and Dynamic time warping distances. We employ these clusters to build short-term load forecasting models using federated learning. Our results reveal the possibility of using P2P clustering along with simple Euclidean distances and FL to obtain highly performant load forecasting models in a fully decentralized setting.

1 Introduction

Load forecasting is one of the most crucial aspects of both traditional and modern power systems (Muñoz et al., (2010)). The main purpose of load forecasting in power system operational planning is to maintain balance between power supply (generation) and demand (load). The increasing penetration of variable renewable energy sources (VRES), electric vehicles, and prosumers (consumers with VRES) challenges the power system balance. They introduce additional volatility and uncertainty leading to higher imbalances and power system operation costs. To maintain the balance, higher accuracy of short-term load forecasting (STLF) models, along with higher accuracy of VRES forecasting models, is necessary (Muñoz et al., (2010)). The STLF models provide forecasts for a time horizon between 1 to 168 hours (Muñoz et al., (2010)).

Traditional STLF models rely on static standard load profiles and only partially capture the variability of the load. Newer data-driven approaches can provide dynamic models that can better capture the variability of the load (Hippert et al., 2001; Nassif et al., 2021).

These data-driven approaches rely on techniques such as Machine learning (ML) and Deep Learning (DL). However, they require a large amount of data and high computing power. The roll-out of smart meters with higher measurement granularity, initiated in many countries in the last couple of years, generates the required amount of data for ML and DL. These ML and DL models can forecast load curves (time series) with higher accuracy compared to traditional methods (Hippert et al., 2001; Nassif et al., 2021; Nti et al., 2020).

Classic ML models, such as Autoregressive integrated moving average (ARIMA) or exponential smoothing, have limiting assumptions, such as linearity. With increasing data granularity, these limitations get accentuated. Modern forecasting techniques, such as DL models, can correctly capture these nonlinear and latent patterns in the data, leading to increases in the accuracy of STLF forecasts. DL models use a wide range of techniques, such as Long short-term memory (LSTM), Convolutional Neural Network (CNN), or even hybrid models that combine multiple neural network architectures. Examples of hybrid models are attention-based methods (Sehovac and Grolinger, 2020), autoencoders (Marino et al., 2016), and deep autoencoders (Kong et al., 2017).

Data expansion and changes in the power system create higher variability among households, and consequently, these variations appear in their load profiles. In this regard, the academic literature has opted to cluster household profiles to increase forecasting accuracy (McLoughlin et al., 2015). Household clusters contain load profiles with similar characteristics. By doing so, the reduced intrinsic variability of the clusters eases the learning of the models (Syed et al., 2021). In turn, fitting the model to a particular dataset and then generalizing it to other clusters would result in high bias and low variance.

In recent years, due to existing data but limited smart meter data access, FL has gained traction as a new framework for STLF as it can overcome these limitations (Fernández et al., 2022; Savi and Olivadese, 2021). FL is a decentralized ML multi-party computation technique that can iteratively and collaboratively train any artificial intelligence (AI) model (McMahan et al., 2017). It provides an alternative to centralized models, as it does not require storing data in a central server (silo) nor exchanges of its peers' (clients') raw data (i.e., smart meter).

STLF can benefit from FL as it reduces the limitations of data availability since peers do not need to share raw data, but rather model parameters (Fernández et al., 2022; Han et al., 2020; He et al., 2021; Savi and Olivadese, 2021). Peer's data present high variability because of their decentralized nature and distinct load consumption patterns. One solution to reduce this variability is to cluster the households based on their load profiles. However, clustering of load profiles usually requires global access to the data (Saxena et al., 2017), i.e. it has centralized structure. This global access opposes to decentralized nature of FL. Consequently, the combination of clustering techniques and FL suffers from incompatibility (Fernández et al., 2022; Han et al., 2020; He et al., 2021; Savi and Olivadese, 2021).

In this study, we propose a P2P decentralized clustering model that allows individual households to collaborate to produce STLF models using FL. We compared our decentralized P2P clustering model to a centralized model commonly used for this purpose. We also included different time series specific distance metrics employed in clustering techniques to generate suitable clusters.

In summary, we propose and evaluate a fully decentralized clustering approach for FL to obtain highly accurate forecasting models. These models could help different energy actors, such as Distribution system operators (DSO) or energy suppliers.

The remainder of this paper is structured as follows. Section 2 provides an overview of different clustering techniques, its distance metrics, and a deeper view of FL. Section 3 presents the clustering logic of the central and peer-to-peer models for later comparison (benchmark). In addition, it provides the FL training process details. Section 4 provides an overview of the evaluation process, covering the evaluation metrics dataset, simulation environment, and procedure. Section 5 compares the results and provides a discussion. Finally, Section 6 provides a conclusion.

2 Background

2.1 Clustering techniques

Clustering algorithms group data into so-called clusters in which elements of the same cluster share similar properties (Saxena et al., 2017). These clustering algorithms can be centralized or decentralized, depending on where data storage and computation occur, and can use supervised or unsupervised learning techniques (Ahuja et al., 2020).

On the one hand, centralized clustering algorithms require central silos to store all the data and a central server to run the clustering algorithm. Most of the academic literature focuses on centralized clustering algorithms (Saxena et al., 2017).

On the other hand, the extension of decentralized clustering is limited. Most of the examples refer to decentralized algorithms evaluated in Agent Based Modeling (ABM) simulations (Xu and Wunsch, 2005). In that way, the clustering algorithms are *ad-hoc* solutions for the particular problem they are solving (Ogston et al., 2003). In ABM, each agent will control a single part of the data set (an agent can be understood as a household), and thus the agents will individually decide on their own. In our case, the agents decide to create or dissolve clusters according to a given similarity metric. Agents can create clusters without needing a central server or silo, enabling a fully decentralized P2P environment and thus moving towards P2P economy (Brazier et al., 2015).

Some methods combine centralized and decentralized characteristics. These methods focus on the decentralization of classic centralized algorithms. For example, Federated k-means (Soliman et al., 2020) can train k-means clustering where distinct clients have shares of the dataset. In Federated k-means the training occurs in rounds where the centroids' moves are averaged every round. In addition, to have meaningful movements, each participant must have a large enough portion of the dataset to replicate the training on the entire data set. This limits the overall scope of the method and requires new fully decentralized methods for FL.

2.2 Distance Techniques

Regardless of the clustering technique, all the clustering algorithms rely on a similarity metric. This similarity can measure statistical correlation between vectors (see metrics such as Pearson's correlation or Spearman's rank correlation) or measure the separation between vectors (distance metrics). Similarity metrics allow the clustering algorithm to estimate whether or not two entries should be in the same cluster.

For this paper, we considered the two leading distance metric approaches to measure the closeness between time series (i.e., household load profiles). These distance metrics are Euclidean and Dynamic time warping (DTW).

2.2.1 Euclidean

A standard metric for comparing two vectors is the Euclidean distance. It requires a point-to-point mapping between comparable observations between two time series. However, in the case of slight misalignment along the time axis (generally the x axis), the distance metric between the two time series becomes significantly affected. Such misalignment can occur due to instrument measurement errors and time delays.

2.2.2 DTW

Under temporal constraints, standard distance measurements, such as Euclidean distances, fail to estimate the similarity of time series. For instance, multiple misalignments and links could simultaneously appear in different phases during the progression of a temporal series.

One solution to time alignment issues that might occur when time series are in phase or at different paces is DTW. Given two sequences X and Y, their DTW distance $D(X, Y)$ is defined as follows:

$$dtw(i,j) = \begin{cases} \infty & \text{if } i = 0 \text{ or } j = 0 \\ 0 & \text{if } i = j = 0 \\ \|X_i, Y_j\| + \min \begin{cases} dtw(i-1,j) \\ dtw(i,j-1) \\ dtw(i-1,j-1) \end{cases} & \begin{matrix} \forall i \in X,\ \forall j \in Y \\ dtw(i,j) \end{matrix} \end{cases} \tag{1}$$

DTW has been widely used to find similarities between time series. However, as seen in (1), DTW is a recursive function over the lengths of the two time series and hence computationally expensive. The computational burden could be bounded to a quasi-quadratic form of $O(n \cdot m)$ where n and m are the lengths of two time series. Although it is a high-complexity computation compared to euclidean methods, it performs well (Gao et al., 2021).

For example, Figure 1 illustrates the visual difference between the DTW and Euclidean distances for the same two household load profiles over 48 hours. DTW finds alignments across the spikes around t = 15 to t = 20, while the Euclidean cannot, as seen in t=42 where the spike in the above profile is measured against a valley in the below profile.

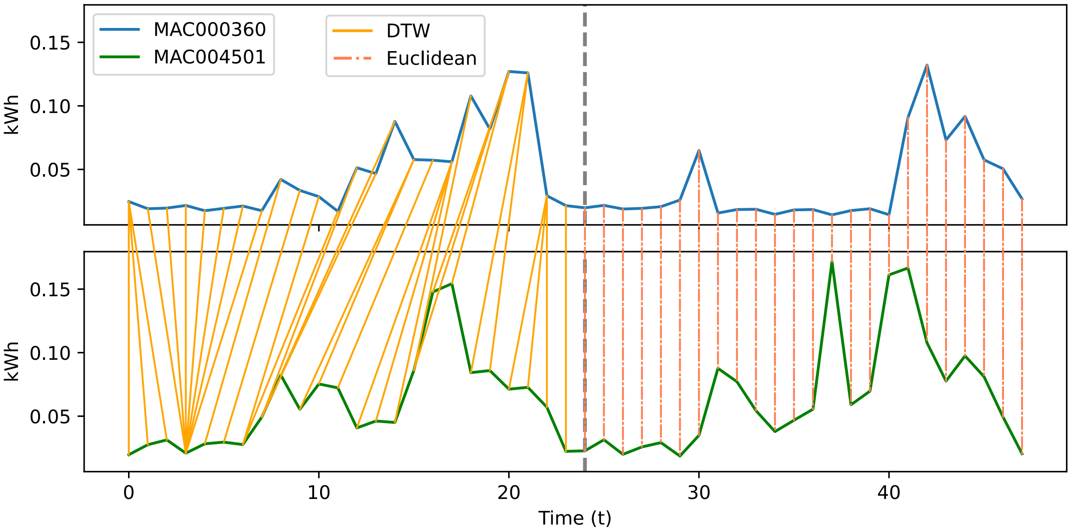


Figure 1: Comparison of DTW (left) and Euclidian (right) distance over 48h for two residential load profiles.

2.3 Federated Learning

FL was introduced in 2016 as a way for entities to train a global model between multiple decentralized clients. Each client uses their local data to train models, without sharing their training data. (McMahan et al., 2017). This shift allowed models to grow by ingesting large amounts of data without the need to store the data in a centralized silo. In FL, clients do not share raw data but information about models, normally with a server. Thus, the server will process the models from the clients, reaching a consensus without accessing the local data of the clients. More specifically, FL works as follows: initially, the server selects an AI model to train. Later, the server shares the model with a random set of clients; normally, the ratio Q represents the fraction of this subset w.r.t. the total number of clients. A higher Q implies that more clients participate in each round of training and vice versa, where $Q = 1$ involves all clients. Once the clients receive the initial model, they begin training it. Each client uses their local data to train the received AI model.

Clients will share different information with the server depending on the FL algorithm. On the one hand, when using the Federated Stochastic gradient descent (Fed-SGD) algorithm, clients share their gradients of the loss at every batch of training. On the other hand, in Federated Average (Fed-Avg), clients share the weights of the trained model after every training epoch. The former requires more communication rounds between the server and the clients (McMahan et al., 2017). Consequently, the latter is often used. The role of the server once it receives information from the clients is straightforward. The server averages the information (gradients or model weights) and shares the averages back to the selected clients for a new training round. By doing so in an iterative manner, clients train and learn from other clients' data without accessing their data. The output is a collaborative model capturing the variability of clients' data.

3 Models

3.1 Decentralized P2P ABM Clustering

We use an ABM approach for our fully decentralized P2P model. For the description of the ABM, we follow the Overview, Design concepts and Details (ODD) protocol (Grimm et al., 2020). We do not consider the ODD+D extension as we do not include human interaction in the model (Grimm et al., 2020).

The models' primary purpose is to demonstrate how households (i.e., agents) can create clusters in a P2P manner. We define only a simple general pattern to assess the model's usefulness: the number of clusters created. It depends on how many agents there are and their interaction.

Our model includes two kinds of entities: smart meters (agents) and federations (collectives). We provide a list of their state variables and their descriptions in Table 1. Within the federation, we refer to f_{ID} as the average distance of the smart meters within a federation, meanwhile f_{EID} refers to the extended average distance of the smart meters within a federation including a new smart meter.

Table 1: Description of entities and their variables.

Entities	Variable	Description
Smart meter	sm_{lp}	Assigned unique load profile
	sm_{list}	A list of calculated distances
	sm_{id}	Individual smart meter ID
	p_s	Asking threshold
Federation	f_{ID}	Internal federation average distance metric
	f_{EID}	Extended federation average distance
	f_{size}	number of smart meters in a federation
	f_{id}	Individual federation ID

We do not correlate time steps to seconds, minutes, or hours; we keep it agnostic. Similarly, the grid is not a physical attribute but an abstract plane. We acknowledge these assumptions as limitations for directly applying our decentralised clustering approach. It would require a to consider the time required per ABM round based on the information communication technology (ICT) properties for smart meter data exchange, the time horizon for a short-term forecast, and the specific area of interest for agents (smart meters) to create clusters. For instance, a DSO seeking local flexibility in a particular secondary substation due to scheduled maintenance may request specific smart meters to create clusters and provide a short-term load forecast, which can aid in acquiring dynamic flexibility.

The procedure that our ABM follows is as such. We select a total number of smart meters from the dataset for our model. Each smart meter will have its unique load profile (see Section 4.3).

We define two subsets. Subset Z is a defined random number of smart meters. Each smart meter of Z has an internal subset Z′, composed by entities. To limit the computation overhead we limit $\|Z'\| = p_s\|Z\|$. Then in an iterative process, each smart meter of Z computes and saves in sm_{list} the distance between itself and each of the entities in Z′.

In the next step, each smart meter of subset Z will sort its distances and choose the shortest distance. Then it can: A) create a federation if none of the smart meters is in a federation, B) join a federation, or C) move to another federation. The last two cases require a deeper explanation. In the case of B), it refers to the smart meter in Z: sm not belonging to a federation, while the smart meter in Z' is in a federation f or viceversa. sm will only join f if f_{EID} is smaller than f_{ID}. In the case of C), both smart meters are part of federations. A smart meter will only change from one federation to another if its movement positively impacts both federation's metrics (f_{ID} and f_{EID}).

The iterative process is repeated for a specified number of rounds. The output of the iterations is an undefined number of federations of different sizes and smart meters which did not join any federation.

To initialize ABM, we only need to choose the total number of smart meters, rounds, and metric distance. In our case, we consider 300 smart meters, 300 rounds, and two possible distance metrics (Euclidian and DTW). Thus, the ABM only requires load profiles as input data. The final output of ABM will be different federations (clusters) as depicted in Figure 2.

3.2 Centralized k-means clustering

K-means is the main representative of centralized clustering. It has been the go-to solution for clustering due to its usefulness and adaptability. K-means is an unsupervised algorithm which randomly initializes a given number of centroids. These centroids are the center of a cluster. In every round, the centroids iteratively move towards the center of mass of each cluster until no more moves are required. K-means has been successfully demonstrated to cluster load profiles, as in (Baliga et al., 2010; Dong et al., 2022), where the authors use k-means to aggregate customer load profiles with high accuracy (Bian et al., 2020).

3.3 Federated Learning

FL requires a baseline learning model. This model could range from simple linear models to AI architectures. We follow the architectural design of (Fernández et al., 2022) and the Artificial Neuronal Network architecture (Marino et al., 2016) to be the baseline of our model. Their model is an encoder-decoder architecture with 12 neurons in the latent space. We collect the hyperparameters of the FL model in Table 2. In particu-

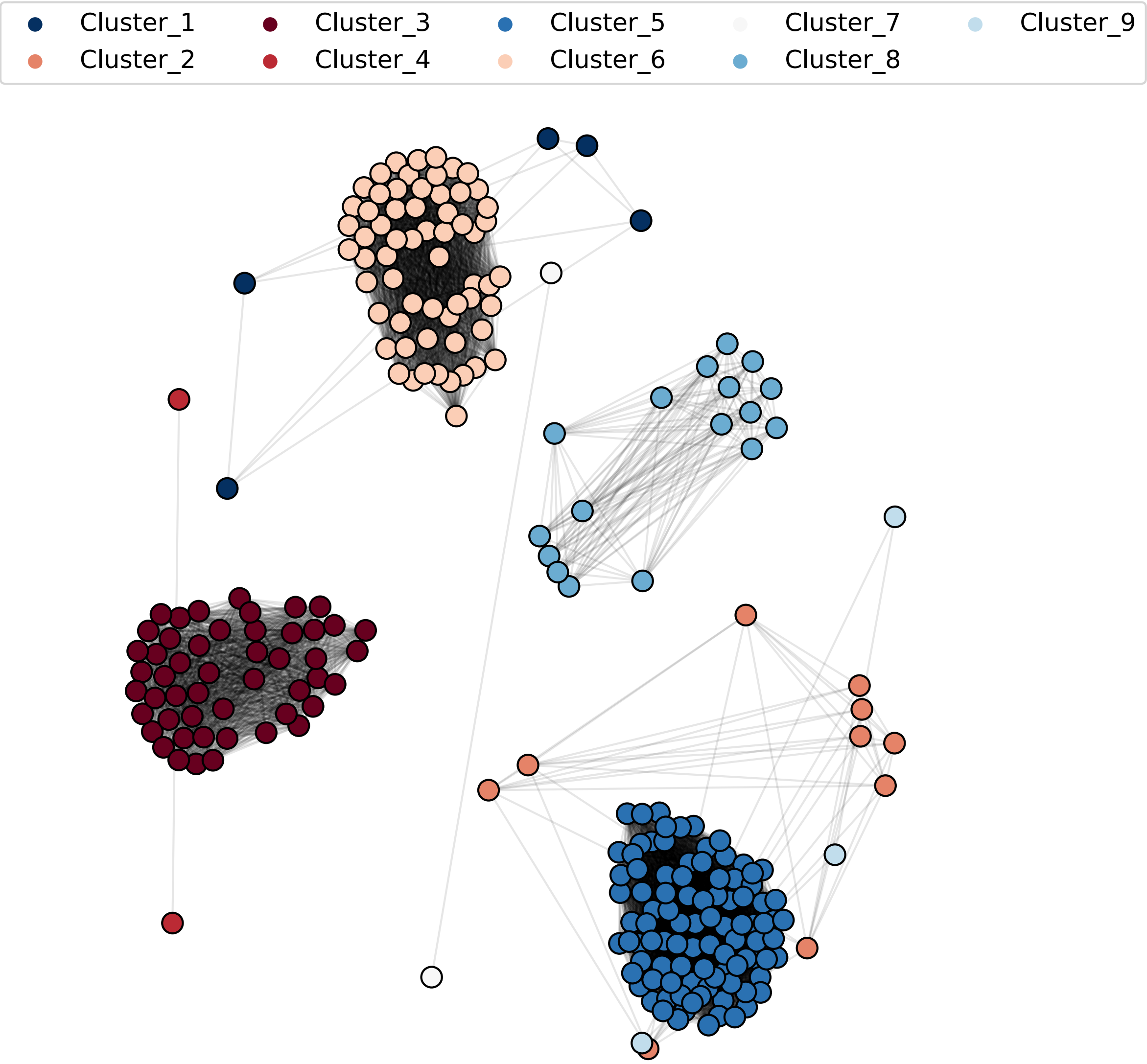


Figure 2: Clusters after 300 rounds using P2P and DTW.

lar, we define Q as a function of the cluster size to limit the computational burden of large clusters. In our case, we define a maximum of 15 clients per round and produce 1-hour-ahead forecasts.

Table 2: Hyperparameters for FL models.

Parameter	**Value**
Number of internal rounds before averaging	5
Artificial Neuronal Network architecture	Marino et al. (Marino et al., 2016)
Clients within a cluster (w)	38
Ratio of clients involved per round (Q)	$Q = \begin{cases} 1.0 & w < 15 \\ \text{w/15} & w \geq 15 \end{cases}$
Optimizer	Adam
Optimizer learning rate (L_r)	10^{-3}
Batch size	128
Number of communication rounds	100

4 Evaluation

4.1 Evaluation metrics

Evaluation metrics offer indicators of the models performance and enable fair comparison. Each of the metrics depicts different characteristics of the models and their ability to predict them. On the one hand, absolute metrics such as MAE (2) or MAPE (3) are known to be robust with respect to outliers. On the other hand, quadratic metrics (MSE (4) and RMSE (5)) penalize large prediction errors, as they measure the standard deviation of residuals.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \tag{2}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| \tag{3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \tag{4}$$

$$RMSE = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(y_i - x_i)^2} \tag{5}$$

4.2 Simulation environment

We performed the simulations in the IRIS Cluster of the high-performance computer (HPC) facilities of the University of Luxembourg (Varrette et al., 2014). The simulations for the clustering ran in an specific node with 1Tb of RAM while the FL model trained on two NVIDIA Tesla V100 with 16Gb or 32Gb depending on the allocation. We programmed FL and the ABM model in Python using Tensorflow-Federated (Authors, 2018) and MESA framework (Kazil et al., 2020) respectively. The DL models were written in Keras (Chollet et al., 2015) and the time series k-means was built using DTAIDistance (Meert et al., 2020).

4.3 Dataset

For our simulations, we used a dataset collected during the Low Carbon London project within the UK Power Networks conducted between November 2011 and February 2014 in the London area (D., 2019). It contains the electrical consumption (kWh) of households in a half-hour resolution. We treated our dataset to be ready for the simulations in the following manner. First, we downscaled the values from half-hour resolution to an hour resolution. This implies a reduction in the computational needs of the models. Second, we drop all the null and outlier values. Third, we rescaled the load profiles to a known range (0 to 1) using Min-Max scaler to further increase the model convergence. Forth, to limit the simulations, we restricted our dataset to 372 profiles. To ensure the validity of our simulations, we split the dataset into training and test set. Initially, we divided our households into two sets. In the first one, we randomly selected 300 households representing the training dataset. In the second one, the remaining 72 we used them to evaluate the performance. Regardless of the previous split, we split each particular household again in training and testing. This split affects the training and test of the FL models; by splitting the data we prevent the model to overfit known patterns and thus evaluate its ability to generalize under new conditions. The training set contains information from January 1st 2013 until December of the same year, while the test set contains data from January 2014 to March 2014.

4.4 Simulation Procedure

To evaluate the performance of both clustering techniques, we established a pipeline as depicted in Figure 3. In step one, we divided the dataset into a training split (300 households) and a test split (76 households). In step two, we perform the clustering. We cluster our data using both P2P and k-means and for each of the clustering techniques, we run one simulation per distance metric (Euclidian and DTW). We optimized the number of k-means clusters using the elbow method (Thorndike, 1953). The results of step two are the clusters. These are the input for the third step. In step three, we trained the FL models based on the clusters and the households inside. The result of this step was as many FL models as clusters found in the previous step. In step four, we estimated the most optimal cluster for each of the households in the test split. From this estimation, we subsequently evaluated the forecasting performance of each FL model.

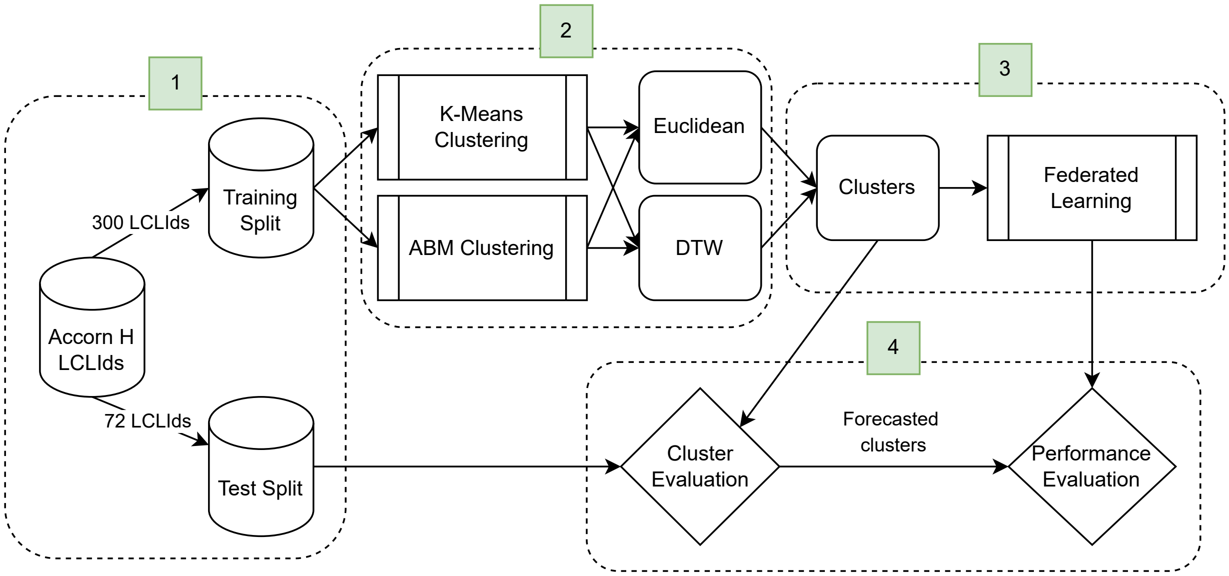


Figure 3: Simulation pipeline.

5 Results

We analyze our results from two points of view. The first is the absolute performance of the FL models based on the chosen metrics (4.1). The second is the computation time of the cluster, calculated in minutes.

We collect in Table 3 the performance of the P2P and k-means clustering results. On average, k-means perform better than P2P. However, the performance difference is insignificant since it is only 0.51 percentage points (pp), and it could be due to the stochastic nature of the models. On the same page, the difference between DTW and Euclidean distances is limited. Their disparities are 0.13 pp, towards Euclidean distances. These results showcase two readable outcomes. First, our P2P clustering approach leads to results similar to those of a central k-means clustering algorithm, facilitating further fully decentralized forecasting approaches. Second, Euclidian and DTW offer similar results. Their similarity might be a consequence of the small shifts in the load profiles of each smart meter present. These small shifts lead to similar measurements across distances and, thus, similar performance.

Concerning the computation time (see Table 3 under T[min]), it diverges dramatically between Euclidean and DTW. Euclidean is on the linear order of $O(Max(n,m))$, while DTW is on the quadratic order of $O(n \cdot m)$, being n and m the lengths of the two input sequences. Our results suggest between 4.5 to 9 times slower to compute the clustering when using the DTW distance metric. This is particularly prominent in our P2P case where the distance computations occur at a much higher rate, thus slowing the convergence of the algorithm by almost double. Our findings imply that applying DTW over Euclidean is not justified for clustering consumer load profiles with similar load profiles (small shifts) given in our case by the ACORN classification.

Table 3: FL performance results of P2P and k-means clustering using Euclidian distances or DTW.

			MAE	MSE	RMSE	MAPE	T[min]
P2P	Euclidean	μ	0.0055	0.0002	0.0122	13.1284	20
		σ	0.0048	0.0005	0.0090	5.7536	-
	DTW	μ	0.0047	0.0002	0.0116	12.3761	180
		σ	0.0044	0.0005	0.0084	6.3897	-
K-means	Euclidean	μ	0.0036	0.0001	0.0102	10.8050	2
		σ	0.0019	0.0002	0.0055	5.4495	-
	DTW	μ	0.0041	0.0002	0.0105	11.8331	90
		σ	0.0029	0.0003	0.0063	5.6228	-

6 Conclusions

Traditionally, the high variability of consumer loads has been tackled by clustering them into similar groups. FL is commonly used with centralized clustering approaches, and even though highly effective, this combination suffers from incompatibilities. This paper proposes P2P decentralized clustering technique to solve these incompatibles. We evaluated a new P2P decentralized clustering technique using ABM and compared it to a k-means approach, a traditional centralized clustering technique. Furthermore, we evaluated two distance metrics for clustering: Euclidian and DTW. Eventually, we trained FL models to predict one-hour-ahead load and analyzed the performance of the forecasts together with the total computation time.

Although we acknowledge that the computational load can be substantial, our simulation ignores the technical details of the processing units and procedures on the smart meter side.

Our decentralized P2P clustering approach produces similar clusters to centralized k-means, even with different distance metrics. The FL models trained for each clustering approach perform similarly. Consequently, the decentralized P2P clustering approach enables fully decentralized FL forecasting models.

Our analysis also suggests that classic Euclidean distances perform similarly to more complicated and slower methods like DTW. Without additional computational burden, Euclidean distances are enough to produce adequate clusters for FL.

Acknowledgements

We would like to acknowledge the support of the European Union (EU) within its Horizon 2020 programme, project MDOT (Medical Device Obligations Taskforce), Grant agreement 814654. Additionally, this project was supported by the Kopernikus-project “SynErgie” of the German Federal Ministry of Education and Research (BMBF), by PayPal and the Luxembourg National Research Fund FNR (P17/IS/13342933/PayPal-FNR/Chair in DFS/Gilbert Fridgen) as well as by the Luxembourg National Research Fund (FNR) – FiReSpARX Project, ref. 14783405.

The experiments presented in this paper were carried out using the HPC facilities of the University of Luxembourg (Varrette et al., 2014)– see hpc.uni.lu

References

Ahuja, R., A. Chug, S. Gupta, P. Ahuja, and S. Kohli (2020). "Classification and Clustering Algorithms of Machine Learning with their Applications". In: *Nature-Inspired Computation in Data Mining and Machine Learning*. Ed. by X.-S. Yang and X.-S. He. Cham: Springer International Publishing, pp. 225–248. ISBN: 978-3-030-28553-1. DOI: 10.1007/978-3-030-28553-1_11. URL: https://doi.org/10.1007/978-3-030-28553-1_11.

Authors, T. F. (2018). *TensorFlow Federated*. https://github.com/tensorflow/federated. Version 0.20.0.

Baliga, J., R. W. Ayre, K. Hinton, and R. S. Tucker (2010). "Green cloud computing: Balancing energy in processing, storage, and transport". In: *Proceedings of the IEEE* 99.1, pp. 149–167.

Bian, H., Y. Zhong, J. Sun, and F. Shi (2020). "Study on power consumption load forecast based on K-means clustering and FCM–BP model". In: *Energy Reports* 6, pp. 693–700.

Brazier, F. M. T., H. la Poutré, A. R. Abhyankar, K. Saxena, S. N. Singh, and K. Tomar (2015). "A review of multi agent based decentralised energy management issues". In: *2015 International Conference on Energy Economics and Environment (ICEEE)*, pp. 1–5.

Chollet, F. et al. (2015). *Keras*. https://keras.io.

D., J.-M. (2019). *Smart meter data from London area*. URL: https://www.kaggle.com/jeanmidev/smart-meters-in-london.

Dong, X., S. Deng, and D. Wang (2022). "A short-term power load forecasting method based on k-means and SVM". In: *Journal of Ambient Intelligence and Humanized Computing* 13.11, pp. 5253–5267.

Fernández, J. D., S. P. Menci, C. M. Lee, A. Rieger, and G. Fridgen (2022). "Privacy-preserving federated learning for residential short-term load forecasting". In: *Applied Energy* 326, p. 119915. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2022.119915. URL: https://www.sciencedirect.com/science/article/pii/S0306261922011722.

Gao, M., S. Pan, S. Chen, Y. Li, N. Pan, D. Pan, and X. Shen (2021). "Identification Method of Electrical Load for Electrical Appliances Based on K-Means ++ and GCN". In: *IEEE Access* 9, pp. 27026–27037.

Grimm, V., S. F. Railsback, C. E. Vincenot, U. Berger, C. Gallagher, D. L. DeAngelis, et al. (2020). "The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism". In: *Journal of Artificial Societies and Social Simulation* 23.2.

Han, F., T. Pu, M. Li, and G. Taylor (2020). "Short-term forecasting of individual residential load based on deep learning and K-means clustering". In: *CSEE Journal of Power and Energy Systems* 7.2, pp. 261–269.

He, Y., F. Luo, G. Ranzi, and W. Kong (2021). "Short-Term Residential Load Forecasting Based on Federated Learning and Load Clustering". In: *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 77–82.

Hippert, H., C. Pedreira, and R. Souza (2001). "Neural networks for short-term load forecasting: a review and evaluation". In: *IEEE Transactions on Power Systems* 16.1, pp. 44–55. DOI: 10.1109/59.910780.

Kazil, J., D. Masad, and A. Crooks (2020). "Utilizing Python for Agent-Based Modeling: The Mesa Framework". In: *Social, Cultural, and Behavioral Modeling*. Ed. by R. Thomson, H. Bisgin, C. Dancy, A. Hyder, and M. Hussain. Cham: Springer International Publishing, pp. 308–317. ISBN: 978-3-030-61255-9.

Kong, W., Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang (2017). "Short-term residential load forecasting based on LSTM recurrent neural network". In: *IEEE Transactions on Smart Grid* 10.1, pp. 841–851.

Marino, D. L., K. Amarasinghe, and M. Manic (2016). "Building energy load forecasting using deep neural networks". In: *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*. IEEE, pp. 7046–7051.

McLoughlin, F., A. Duffy, and M. Conlon (2015). "A clustering approach to domestic electricity load profile characterisation using smart metering data". In: *Applied energy* 141, pp. 190–199.

McMahan, B., E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas (2017). "Communication-efficient learning of deep networks from decentralized data". In: *Artificial intelligence and statistics*. PMLR, pp. 1273–1282.

Meert, W., K. Hendrickx, T. Van Craenendonck, and P. Robberechts (2020). *DTAIDistance*. Version v2.3.10. DOI: 10.5281/zenodo.7158824. URL: https://doi.org/10.5281/zenodo.7158824.

Muñoz, A., E. F. Sánchez-Úbeda, A. Cruz, and J. Marín ((2010)). “Short-term Forecasting in Power Systems: A Guided Tour”. In: *Handbook of Power Systems II*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 129–160. ISBN: 978-3-642-12686-4. DOI: 10.1007/978-3-642-12686-4_5. URL: https://doi.org/10.1007/978-3-642-12686-4_5.

Nassif, A. B., B. Soudan, M. Azzeh, I. Attilli, and O. AlMulla (2021). “Artificial Intelligence and Statistical Techniques in Short-Term Load Forecasting: A Review”. In: *ArXiv* abs/2201.00437.

Nti, I. K., M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya (2020). “Electricity load forecasting: a systematic review”. In: *Journal of Electrical Systems and Information Technology* 7.1, p. 13. DOI: 10.1186/s43067-020-00021-8. URL: https://doi.org/10.1186/s43067-020-00021-8.

Ogston, E., B. Overeinder, M. van Steen, and F. Brazier (2003). “A Method for Decentralized Clustering in Large Multi-Agent Systems”. In: *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*. AAMAS ’03. Melbourne, Australia: Association for Computing Machinery, 789–796. ISBN: 1581136838. DOI: 10.1145/860575.860702. URL: https://doi.org/10.1145/860575.860702.

Savi, M. and F. Olivadese (2021). “Short-Term Energy Consumption Forecasting at the Edge: A Federated Learning Approach”. In: *IEEE Access* 9, pp. 95949–95969. DOI: 10.1109/ACCESS.2021.3094089.

Saxena, A., M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, et al. (2017). “A review of clustering techniques and developments”. In: *Neurocomputing* 267, pp. 664–681. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2017.06.053. URL: https://www.sciencedirect.com/science/article/pii/S0925231217311815.

Sehovac, L. and K. Grolinger (2020). “Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks With Attention”. In: *IEEE Access* 8, pp. 36411–36426. DOI: 10.1109/ACCESS.2020.2975738.

Soliman, A., S. Girdzijauskas, M.-R. Bouguelia, S. Pashami, and S. Nowaczyk (2020). “Decentralized and adaptive k-means clustering for non-iid data using hyperloglog counters”. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pp. 343–355.

Syed, D., H. Abu-Rub, A. Ghrayeb, S. S. Refaat, M. Houchati, O. Bouhali, and S. Bañales (2021). “Deep Learning-Based Short-Term Load Forecasting Approach in Smart Grid With Clustering and Consumption Pattern Recognition”. In: *IEEE Access* 9, pp. 54992–55008.

Thorndike, R. L. (1953). “Who belongs in the family”. In: *Psychometrika*. Citeseer.

Varrette, S., P. Bouvry, H. Cartiaux, and F. Georgatos (2014). “Management of an Academic HPC Cluster: The UL Experience”. In: *Proc. of the 2014 Intl. Conf. on High Performance Computing & Simulation (HPCS 2014)*. Bologna, Italy: IEEE, pp. 959–967.

Xu, R. and D. C. Wunsch (2005). “Survey of clustering algorithms”. In: *IEEE Transactions on Neural Networks* 16, pp. 645–678.

Research Paper 3 – *Federated Learning for Credit Risk Assessment*

Full Title:
Federated Learning for Credit Risk Assessment

Authors:
Chul Min Lee, Joaquín Delgado Fernández, Sergio Potenciano Menci, Alexander Rieger, Gilbert Fridgen

Published in: Hawaii International Conference in System Sciences (HICSS) (2023)

Keywords:
Federated Learning, Artificial Intelligence, Credit Risk Assessment, Financial Collaboration

DOI:
https://hdl.handle.net/10125/102676

Abstract

Credit risk assessment is a standard procedure for financial institutions (FIs) when estimating their credit risk exposure. It involves the gathering and processing quantitative and qualitative datasets to estimate whether an individual or entity will be able to make future required payments. To ensure effective processing of this data, FIs increasingly use machine learning methods. Large FIs often have more powerful models as they can access larger datasets. In this paper, we present a Federated Learning prototype that allows smaller FIs to compete by training in a cooperative fashion a machine learning model which combines key data derived from several smaller datasets. We test our prototype on an historical mortgage dataset and empirically demonstrate the benefits of Federated Learning for smaller FIs. We conclude that smaller FIs can expect a significant performance increase in their credit risk assessment models by using collaborative machine learning.

1 Introduction

Most financial institutions (FIs) employ comprehensive credit risk models to estimate their exposure to credit risk. These models typically employ either traditional or advanced methods. Traditional methods rely on induction principles to make mathematical and statistical inferences from curated data. They facilitate the creation of static models that build on a range of assumptions, such as linearity, independence, and normality. Advanced methods, in turn, are more data-driven and less reliant on these assumptions (Chen et al., 2016; Galindo and Tamayo, 2000). Like traditional methods, they infer information from curated data but they enable the creation of flexible models that adapt to the curated data. As a result, credit risk models that employ advanced methods typically perform better at to extracting patterns from complex real-world datasets that are replete with noise, nonlinearity, and idiosyncrasies.

Both methods depend strongly on data inputs (Altman, 2002; Heitfield, 2009). Models trained with more and better data can estimate real word situations more accurately. In effect, data availability is crucial for FIs and can translate into a competitive advantage (Bansal et al., 1993; Walczak, 2001). Limited data, in turn, can lead to less reliable predictions. For smaller FIs with limited data access, this effectively means that 'data sharing' with other FIs could have a material impact on the performance of their credit risk models (Bansal et al., 1993; Walczak, 2001). However, data sharing is often challenging due to concerns about privacy, control and legal recourse (Borgman, 2012; Ekbia et al., 2015).

A more feasible alternative could be the use of Federated Learning (FL) to create joint credit risk models. FL is an ML technique that allows models to train on a distributed basis without the need to move raw data (McMahan et al., 2016). In other words, financial institutions would not need to reveal their data as they gain insights from its processing, allowing every participating FI to benefit from use of each other's information.

In this paper, we thus ask the following two research questions:

1. How does FL based credit risk assessments perform?
2. Will FL help to reduce the disparities in risk calculations between financial institutions?

To answer these research questions, we developed an FL-based credit risk assessment prototype. We tested our prototype on Freddie Mac's Single Family Loan-Level Dataset (Freddie Mac, 2021b) to simulate collaboration between FIs when assessing the credit risk of mortgage portfolios. Mortgages are an important financial instrument, but their typically long time horizons complicate the task of making accurate forecasts. Specifically, we compared the performance of credit risk models under different scenarios to evaluate and quantify the impact of information sharing. These comparisons indicate that FL can offer significant performance gains for smaller FIs with limited in-house datasets. To the best of our knowledge, this paper is the first to examine FL in assessing the credit risk of mortgages with real-world FI divisions.

The research paper is structured as follows. Section 2 provides an overview of relevant literature on credit risk assessment and federated learning. Moreover, it presents previous research that studies the application of FL in financial services. Section 3 describes the implementation of our FL prototype. Section 4 details the hypotheses, scenarios, and evaluation metrics we used to examine the performance of our FL prototype. Section 5 presents the results of our evaluation. Section 6 discusses the limitations of our study as well as future research directions. Section 7 offers concluding remarks.

2 Related Work

2.1 Credit Risk Assessment

Credit risk assessment methods have evolved over time from traditional to advanced methods, but essentially start and end in the same fashion. At the start, data or information about the prospective mortgage is gathered systematically. Subsequently, the newly collected information is used to measure the likelihood of the mortgage to experience credit risk events. The likelihood of these events results in a score representing the credit risk of the mortgage.

Performance of credit risk models is not only dependent on the method used but also on data inputs (Altman, 2002; Heitfield, 2009). Models trained on larger datasets, for example, when sharing data, allow for more real world data to be represented in the trained model. As a result, changes to the quantity and quality of data inputs have material impacts on the performance of credit risk models.

Regulators and policymakers have called for increased disclosure of credit risk related data (Banking Supervision, 2018) and developed infrastructure to encourage voluntary sharing (Bank, 2010; Israël et al., 2017). However, concerns about sharing data remain and are two-fold. Firstly, data privacy laws, such as the EU's General Data Protection, prohibit data sharing without an appropriate legal basis. Secondly, data typically offers a competitive advantage to its holder (Kearns and Lederer, 2004; Redman, 1995; Zuiderwijk et al., 2015). Therefore, companies are often reluctant to share their data to avoid risking the disclosure of valuable information.

2.2 Federated Learning

FL was introduced as a collaborative ML technique in (McMahan et al., 2016; McMahan et al., 2017) and might help to mitigate privacy and competitiveness concerns. In FL, data remains decentralized across collaborating clients. These clients collaborate through share information (or inferences) about the data rather than the data itself. FL typically builds on one of two algorithms: Federated Averaging (Fed-Avg) and Federated Stochastic Gradient Descent (Fed-SGD). The first algorithm shares model while the second model gradients.

In FL (McMahan et al., 2016; McMahan et al., 2017), there are typically two roles: clients and the central server. The roles are the same for both Fed-Avg and Fed-SGD. Clients host and locally compute ML models locally using their own data. The central server coordinates the sharing of the locally computed information from clients by aggregating, averaging and then distributing the averaged information back to the collaborating clients.

There are four FL variants in standard practice, with these based on the data structures and features used to train the FL models: Horizontal, Vertical, Transfer Learning, and Assisted Learning. Horizontal FL requires that the data used to train each client have the same data structure and features. Vertical FL requires that each client has the same structure but different data features. Transfer learning allows each client to have different structures and features in their data. Assisted Learning allows each client to train using other clients' errors.

The training of an ML model with FL follows an iterative process as depicted in Figure 1. The training steps are as follows: initially, in (1), the central server selects a list of collaborating clients and an ML model to be run by each of the selected clients. Subsequently, in (2), the central server communicates the selected ML model to each randomly selected client. After receiving the selected ML model, in (3), each client simultaneously trains the selected ML model on their data and produces a newly trained model. In (4), each client communicates to the central server the computed results of their new ML models. Once the central server has aggregated the information from all clients, in (5), it will average the aggregated information. Lastly, in (6) the aggregated information is relayed back to each newly randomly selected subset of clients. This collaborative training process continues repetitively between steps (2) and (6) until a prescribed number of rounds are complete, or a target goal is reached.

2.3 Federated Learning for Financial Services

FL is a relatively novel ML method that allows disconnected entities to train ML models without sharing their raw data. At the same time, there is a burgeoning quantity of literature in various fields which study the potential impact of FL on analytical capabilities, such as medical imaging (Kaissis et al., 2021), the Internet of Things (Aïvodji et al.,

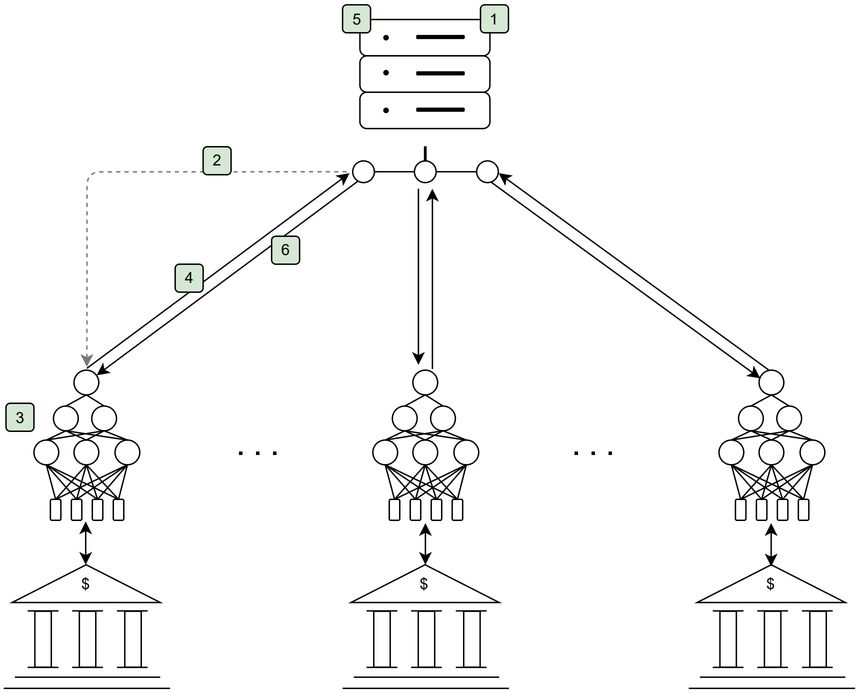


Figure 1: Diagram of Federated Learning process.

2019), and energy demand optimization (Saputra et al., 2019), which study the potential impact of FL on analytical capabilities.

FL is also gaining traction amongst financial services businesses. For instance, Yang et al. (2019) proposes a FL framework to train fraud detection models. They use an anonymized real-world dataset of credit card transactions from European cardholders provided by the Université Libre de Bruxelles (ULB) ML Group. Their framework demonstrates an increase in performance of approximately 10% when implementing FL versus conventional ML approaches. Zheng et al. (2020) propose an FL framework to train meta-learning based models. They test their proposed framework on four publicly available credit card transaction datasets. These tests demonstrate an increase in performance compared to conventional meta-learning approaches. Shingi (2020) applies an FL model to predict loan defaults using a modified learning algorithm for FL and a Feed-Forward Network exhibiting a 3.88% increase in performance. However, the current literature still lacks real portfolio divisions of small, medium, and large FI to create credit risk models that can consume large datasets. Thus, we contribute to reducing this gap in our paper.

3 Prototype

To evaluate empirically the effectiveness of FL in assessing the credit risk of mortgages, we developed an FL prototype. The developed FL prototype estimates the probability of default of the underlying mortgages. We use historical data of mortgage transactions and real division of entities for our federated clients.

3.1 Data Source

The datasets used to train and test our FL prototype are Freddie Mac's Single-Family Loan-Level (FMSFLL) (Freddie Mac, 2021b), Freddie Mac's House Price Index (FMHPI) (Freddie Mac, 2021a), United States Bureau of Labor Statistics' Local Area Unemployment Statistics (LAUS) (United States Bureau of Labor Statistics, 2021), and Federal Reserve Economic Data (FRED) (Federal Reserve, 2021). While the FMSFLL dataset provides data directly related directly to mortgage transactions, the FMHPI, LAUS, and FRED datasets provide complementary data related to economic and environmental factors.

The FMSFLL dataset holds historical records of credit performance data on all mortgages that Freddie Mac has purchased or guaranteed since 1999 and covers approximately 45.5 million mortgages. The dataset has two tables: origination and monthly. The origination table has 35 variables and includes data relevant to when the FI granted the mortgage to the applicant. The monthly table concerns data relevant to the status of the mortgage granted at monthly intervals. The "Seller Name" describes the originating financial institution that initially funded the mortgage transaction at its inception. It allowed us to isolate the mortgages that originated from each FI as true portfolio holdings before Freddie Mac acquired them.

We complemented the FMSFLL dataset with the FMHPI, LAUS, and FRED datasets to include relevant 'environmental' factors for mortgage defaults. Intuitively, the factors considered are general levels of housing price, unemployment, delinquency, charge-off, and interest rates. The FMHPI dataset holds historical housing price levels in the US by state. The LAUS dataset holds historical unemployment levels in the US by state. Lastly, the FRED dataset holds delinquency, charge-off, and interest rate levels in the

US at the national level. As a result, the FL prototype considers descriptive information about mortgages and relevant ’environmental’ factors for mortgage defaults.

3.2 Data Pre-processing

We pre-processed our combined dataset to make it more accessible for our prototype. Firstly, due to the large amount of variables, we reduced the number of those we used to 31 (Table 1). Secondly, to reduce the scope of our evaluation, we worked only with data points from 2006 until 2009. We chose this time frame because the US mortgage markets had a high rate of defaults in those years, which provided an ideal period to test our prototype. Thirdly, to have a consistent terminating state, we only considered mortgage records in the FMSFLL dataset which had default and non-default "termination events" in the Zero Balance Code variable. This only includes mortgage records that experienced credit events such as "Third Party Sale", "Short Sale or Charge Off", "Repurchase prior to Property Disposition", and "REO disposition".

After pre-processing, the combined dataset resulted in 9.6M observations from 250k unique mortgages previously held by 14 FIs.

3.3 Prototype Implementation

The combined dataset after pre-processing is time-stamped at monthly intervals. As time-stamped datasets allow for the consideration of temporal patterns, we chose a Neural Network (NN) with four layers of Long Short-Term Memory (Hochreiter and Schmidhuber, 1997) interspersed with dropout layers and two fully connected layers. Unlike fully connected layers, LSTM layers have feedback connections with previous neurons. These connections allow neurons to access information about their former states so they can make inferences about the future based on previous data.

In NNs, the frequent use of dropout layers mitigates overfitting (Srivastava et al., 2014). During the training phase, dropout will deactivate some neurons at random, encouraging the network to find ways around previously established patterns and preventing some neurons from becoming a bottleneck in the architecture. As a result, the selected architecture can find temporal patterns in the development of the mortgage market (Sezer et al., 2020).

Table 1: List of variables.

Variable	**Dataset**	**Data-type**
Channel	FMSLL	Discrete
Charge-Off Rate	FRED	Continuous
Combined Unemployment Rate	LAUS	Continuous
Credit Score	FMSLL	Continuous
Current Actual Unpaid principal balance	FMSLL	Continuous
Current Loan Delinquency Status	FMSLL	Discrete
Delinquency Due to Disaster	FMSLL	Discrete
Delinquency Rate	FRED	Continuous
Estimated Loan-to-Value (ELTV)	FMSLL	Continuous
First Time Homebuyer Flag	FMSLL	Discrete
Fixed Rate Mortgage Average	FRED	Continuous
House Price Index	FMHPI	Continuous
Loan Age	FMSLL	Continuous
Loan Purpose	FMSLL	Discrete
Loan Sequence Number	FMSLL	Discrete
Mortgage Insurance Percentage (MI %)	FMSLL	Continuous
Number of Borrowers	FMSLL	Continuous
Number of Units	FMSLL	Continuous
Occupancy Status	FMSLL	Discrete
Original Debt-to-Income (DTI) Ratio	FMSLL	Continuous
Original Interest Rate	FMSLL	Continuous
Original Loan Term	FMSLL	Continuous
Original Loan-to-Value (LTV)	FMSLL	Continuous
Property State	FMSLL	Discrete
Property Type	FMSLL	Discrete
Property Valuation Method	FMSLL	Discrete
Remaining Months to Legal Maturity	FMSLL	Continuous
Seller Name	FMSLL	Discrete
Super Conforming Flag	FMSLL	Discrete
Unemployment at origination	FRED	Continuous
Zero Balance Code	FMSLL	Discrete

We implemented the FL prototype using the algorithm presented in (McMahan et al., 2017). In a first step, the central server initializes the baseline model and distributes it to the FIs. In a second step, the FIs start training the model sent by the central server on their local data. In the training they conduct, they use Stochastic Gradient Descend (SGD) (Robbins and Monro, 1951) as the model optimizer with the parameters defined in Table 2. As we are implementing Fed-Avg, the communication rounds happen after one or more complete pass through the dataset, We found 10 internal rounds before averaging the optimal number of rounds. Once all the FIs have finished their training, they share their models weights' with the central server. The central server averages the weights, creating a new model. This model is then shared again with the FIs until a pre-determined number of rounds is reached. In our case, after 100 rounds, there was no improvement in the performance metrics.

We simulated all the FIs and the communications on the High-Performance Computing facilities of the University of Luxembourg's (Varrette et al., 2014). The hardware used was 256Gb of RAM and one 16Gb/32Gb NVIDIA Tesla V100 depending on the allocation. We implemented the FL architecture in TensorFlow Federated *TensorFlow Federated* (2018) while for the Deep Learning modules we used Keras (Chollet et al., 2015)

Table 2: Hyperparameters for FL models.

Parameter	**Value**
Rounds before averaging	10
Baseline architecture	4x (LSTM + Dropout) + 2 Dense
Total number of FIs	14
Optimizer	SGD
Optimizer Learning rate (L_r)	0.01
Optimizer Momentum (v_{t+1})	0.9
Optimizer Decay (λ)	1e-2/100
Batch size	128
Number of communication rounds	100

4 Evaluation

To analyze the performance effects of the FL prototype, we formulated a null and an alternative hypothesis, defined metrics to measure performance, and designed a series of scenarios to test the hypotheses.

4.1 Hypotheses

We formulated our null and alternative hypotheses as follows: Given an FI's dataset F_i: $\{F_1, F_2, ..., F_n\}$ and a global dataset $D = \{F_1 \cup F_2 \cup ... \cup F_n\}$ with n being each individual FI, the null hypothesis (H_0) is that the performance of models trained collaboratively through FL on F_i is better than the performance of the same model trained on F_i. The alternative hypothesis (H_1), in turn, is that the performance of models trained collaboratively through FL is worse than the same model trained on F_i.

4.2 Evaluation Metrics

We used a range of standard metrics to measure the performance levels of the models: Accuracy, Recall, Precision, and F1. The performance of any classification task could be summarized using four main indicators: True positive (TP) representing the instances correctly classified, similarly true Negatives (TP) where the model correctly predicts the negative class. On the other hand false positive (FP) represents the instances incorrectly predicted as positive class. False negative (FN) represents the instances incorrectly predicted as negative class. Eq. 1 Accuracy describes the proportion of correct predictions as opposed to the total number of predictions. Eq. 2 Recall describes the proportion of positive classifications that were correctly classified over the all the positive instances Eq. 3 Precision describes the proportion of positive classifications that were correctly classified among all instances. Eq. 4 F1 is the equal weighted harmonic average of Eq. 2 and Eq. 3.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

4.3 Evaluation Scenarios

We designed a series of five scenarios to test the null hypothesis: 1) Local Model, 2) Central Model , 3a) FL, 3.b) FL without the biggest FI (n-1), and 3.c) FL without the two biggest FIs (n-2). Each scenario represents a hypothetical instance of credit risk assessment. Each scenario uses data from the years 2006 to 2008 as training data, and data from year 2009 as testing data. We calculated the performance metrics for the five scenarios in relation to the observed loan termination status.

Table 3: Scenario and the data they ingest.

Scenario Name	Train data	Tested data
1. Local Model	Own	Own
2. Central	All	All
3a. FL	FL	Local
3b. FL n-1	FL	Local
3c. FL n-2	FL	Local

1. **Local Model** This scenario explores independent FIs that only use the data at their immediate disposal and without collaboration for credit risk assessment. It denotes an 'imperfect information' scenario in which FIs do not have access to each other's data. In this scenario, we trained one model for each FI.

2. **Central** This scenario represents a hypothetical data lake in which all data is pooled together in a singular or centralized data silo for credit risk assessment. It represents a 'perfect information' scenario in which all the data of every FI is available. In this scenario, we trained one model in a 'centralized' manner.

3. a) **Federated Learning** This scenario represents collaboration using Horizontal FL. Each FI stores their own data while their data structure remains identical among clients.

   b) **Federated Learning without the biggest bank** This scenario explores collaboration without Wells Fargo Bank, N.A. Wells Fargo is the FI with the largest number of unique mortgages in our dataset. It holds 40.21% of the total number of unique mortgages.

c) **Federated Learning without the two biggest FIs** This scenario is similar to the previous one and explores the impact of removing the two biggest FIs. Wells Fargo and Chase Home Finance are the two FIs with the largest number of unique mortgages. The two account for 52.59% of the total number of unique mortgages.

To ensure that the results are robust and to normalize the effects of NN's random nature, we utilized a Monte Carlo simulation (Kroese et al., 2014). The n in Table 4 presents the number of simulations for every particular scenario. Additionally, we summarized the metrics by the mean μ and their standard deviation σ. In the **Local Model** scenario, we compute μ and σ across the different FIs for each Monte Carlo simulation.

5 Results

Based on our simulations, we fail to reject the H_0 (the FL model is better than the local model). The hypothesis holds even for scenarios where the largest and two largest FIs do not collaborate in training a forecasting model to predict credit risk. To support our rejection, we provide our simulation results in Table 4.

We found that the local model results offer an average performance worse than the other models. Moreover, we found that the performance of the models was proportional to the number of mortgages on which they were trained. To quantify this effect, we fitted a linear regression between the number of mortgage records and the performance of the models. We found that a 1% increase in the number of loans increased the performance by an average of 0.06% with $p = 0.02 \leq 0.05$.

This relationship between data quantity and model performance explains the variability in the evaluation metrics. For example, the model for Metlife Home Loans, a division of Metlife Bank, N.A., that holds 45340 mortgage observations (0.33% of the total number of observations in our dataset), had 91.12% recall, 75.97% accuracy, a F1 score of 69.44%, and 56.56% precision. Meanwhile, the Wells Fargo Bank NA model with almost 4m mortgage observations (40.20% of the total observations) had 98.97%, 97.95%, 97.35%, and 97.65% accuracy, precision, recall, and F1 score, respectively. In effect, we can con-

Table 4: Performance comparison between scenarios.

Scenario	Accuracy	Recall	Precision	F1 score
1. Local Model	μ = 95.04%	96.97%	89.76%	92.65%
$n = 10$	σ = 0.0667	0.0249	0.1359	0.0879
2. Central	μ = 98.59%	99.8%	95.56%	97.49%
$n = 10$	σ = 0.0263	0.0041	0.0845	0.0468
3a. FL	μ = 99.06%	98.81%	97.69%	98.25%
$n = 10$	σ = 0.0002	0.0005	0.0008	0.0004
3b. FL n-1	μ = 99.04%	98.74%	97.69%	98.21%
$n = 10$	σ = 0.0002	0.0006	0.0011	0.0005
3c. FL n-2	μ = 99.04%	98.72%	96.82%	98.22%
$n = 10$	σ = 0.0004	0.0007	0.0015	0.0007

clude that the higher the number of records per financial institution, the higher their performance levels.

On the contrary, the central model created by all FIs sharing their data in a silo outperforms the **Local Model** scenario by a relatively average performance increase of 4.27 percentage points (pp). However, the difference is not significant between the central and FL models. The difference is 0.57 pp in favor of the FL model; this difference being negligible mainly due to the stochasticity of the models.

Surprisingly, even when we excluded the FIs with the most mortgage records (**FL n-1** and **FL n-2**), the performance levels still matched those of the **Central** and **FL** scenario. Even without the 40.21% and 52.59% respectively of the total mortgage observations, performance did not drop significantly.

The standard deviation over our Monte Carlo simulations (see Table 4 under σ) indicates that there are only minor variations across simulations, exemplifying the robustness of the results. These findings demonstrate that small-to-medium FIs could significantly improve their credit risk assessments by joining forces with others to create a collaborative FL model.

Overall, each FI holds different data. This variation in data induces each FI to estimate credit risk differently, sometimes creating overexposure and other times underexposure. First, we explored how these differences between models develop over time. Thus, as an example, we considered two different loans, F09Q10036282 and F09Q10037931, and explore how different models estimate their risk throughout the life cycle of the mortgage. We illustrate the results of these two loans in Figure 2, where the former defaulted (upper), whereas the latter did not (lower). For instance, we can observe in the predictions for F09Q10036282 that even though all models correctly estimate its default at the end, during the middle years, the default estimations vary across FIs. Even in these two instances, in Figure 2, the difference between the risk estimation between **Central** and **FL** remains negligible.

Secondly, we measured these differences in risk estimation over time. To do so, we calculated the kernel density approximations (Rosenblatt, 1956) of the differences in risk estimation. Individual FIs do not have a complete view of the market and tend to perform poorly in estimating risk distributions. As a complementary step, we add Figure 3, which visualizes how the **Local Model** estimates risk compared to both **Central** and **FL** models. FIs deviate from both **Central** and **FL** when calculating their risk. For example, Metlife Home Loans, a Division of Metlife Bank, N.A., tends to underestimate risk (-25%). Another example is Chase Home Finance LLC which overestimates by around 7%.

Furthermore, Table 5 collects the simple average of the default probability deviations and complements Figure 3. In simple average terms, the **Local Model** scenario underestimates the probability of default compared to the **Central Model** and **FL** scenario at both the initial (2.7% and 5.6%) and final month of mortgages (6% and 4.3%). These results can be explained since a single institution's data has less variability than the rest of the datasets combined. For instance, individual FIs seem to better estimate risk at the beginning of the mortgage but underestimate the default probability at the end. These results are reasonable since an accurate initial default estimation may be simpler to make than analyzing the changing environmental and macroeconomic conditions mortgages are subject to. Hence, FIs could be failing to estimate the fat tails of the mar-

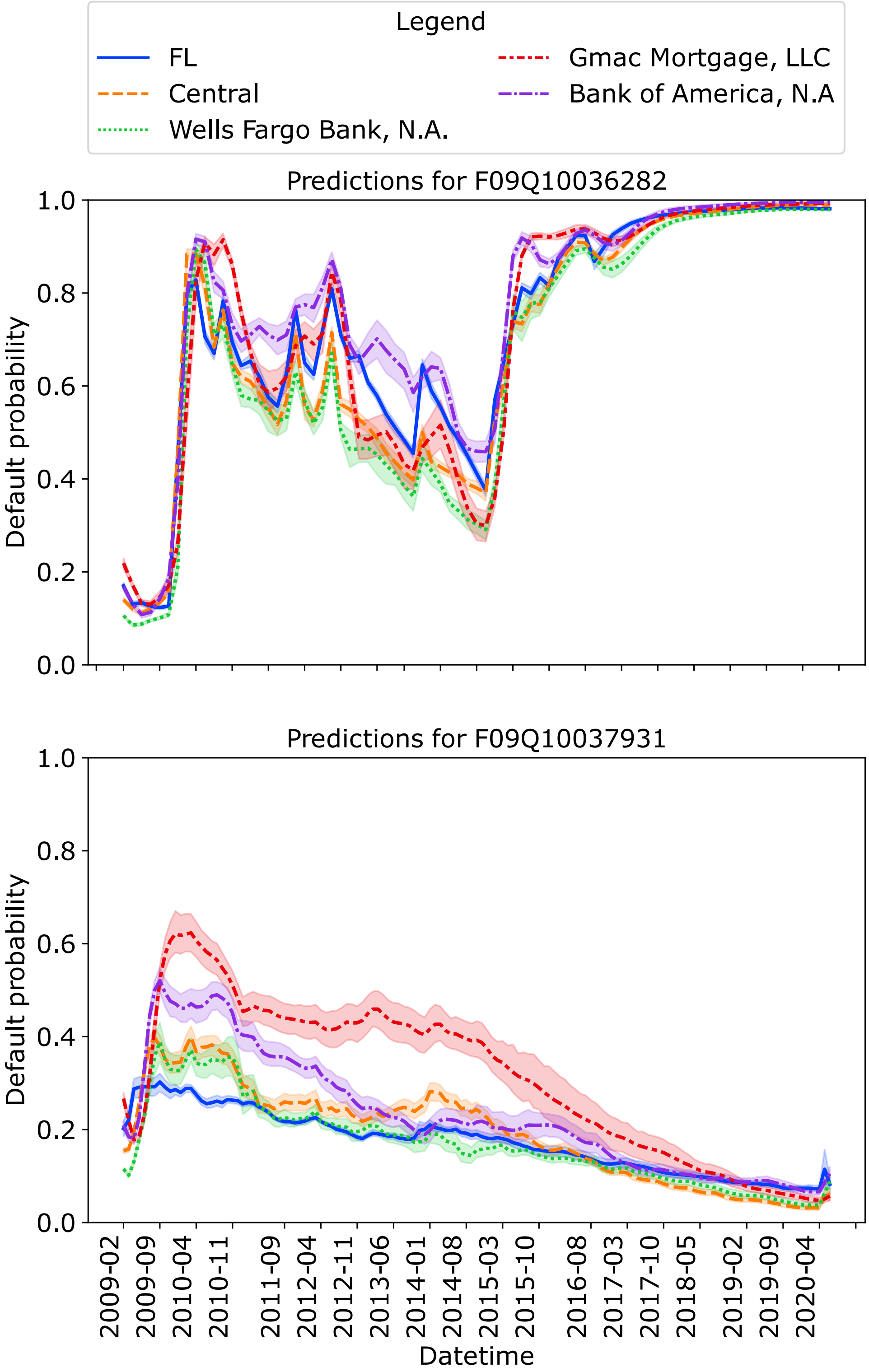


Figure 2: Different models used to estimate default probability.

ket's risk distribution (Taleb, 2020). In other words, smaller FIs may not have access to a *big-picture* view of the risk distribution.

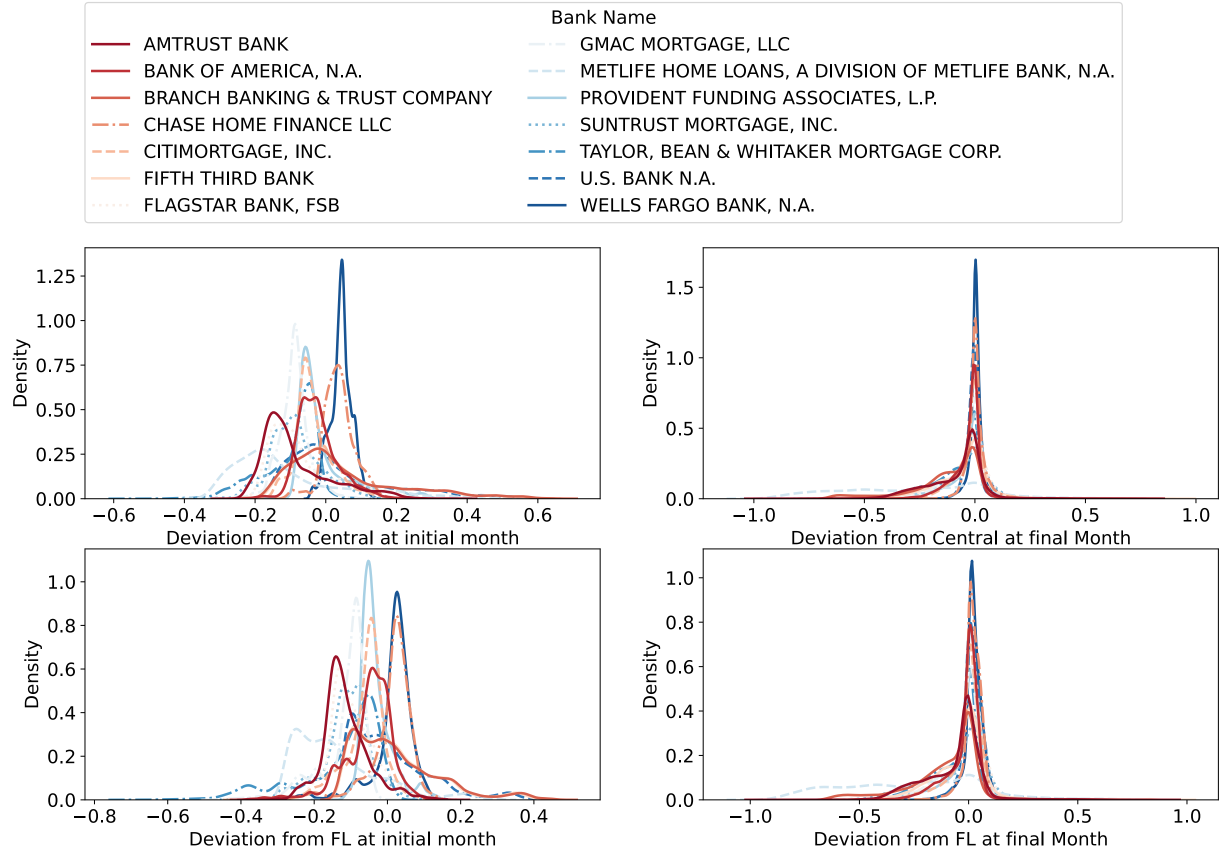


Figure 3: Relative deviation by the local models compared with central and FL scenarios. On the right, the comparison is between FL and Central.

Table 5: Simple average, standard deviation, min and max values for Figure 3. For the local models, the average is across mortgages and then across FIs.

Deviations	μ_s	σ	**min**	**max**
Central at initial month	-0.027	0.119	-0.579	0.639
Central at final Month	-0.060	0.153	-0.980	0.929
FL at initial month	-0.056	0.104	-0.710	0.465
FL at final Month	-0.043	0.156	-0.968	0.971
FL to Central at initial month	0.028	0.061	-0.117	0.283
FL to Central at final month	-0.017	0.048	-0.474	0.765

In essence, there are significant benefits to collaboration through FL for smaller FIs. FIs whose datasets are large enough to approximate the overall variability in the market, in turn, do not significantly benefit from collaboration. Finally, we observe different estimations of default probability by each FIs. These differences vary during a mortgage. Therefore, each FI should individually assess the benefits of applying FL to the credit risk assessment of mortgages.

6 Limitations and Future Research

Our study is subject to two limitations: 1) in that we used only a subset of our dataset; 2) in that we only used the FMSFLL holding division; and 3) in that we worked only with mortgages with final status.

While the full FMSFLL dataset contains approximately 45 million unique mortgages spanning from 1999 to 2022, we used only those 250k that where active from 2006 to 2009. We focused on these years as they had a high number of defaults (Murphy, 2008). Time frames with less defaults, in turn, might lead to different results for the five scenarios. Further research should thus extend our study also to such other time frames.

Moreover, the FMSFLL dataset only includes mortgages that Freddie Mac has bought. In reality, the US FIs' mortage portfolio holdings contains more mortgages than just the ones Freddie Mac bought. We used the FMSFLL dataset because Freddie Mac is one of the major players in the US residential mortgage market. Furthermore, the portfolio holding divisions by each US FI in the sample subset are not arbitrary or random but based on true values and reflect mortgages that each US FI originated or had once held. However, the study could improve by including a hypothetical yet more realistic representation of each FI's mortgage portfolio holdings.

In addition, due to the high volume of loans originated over these years, we limited the number of mortgages by filtering out those with a final status so that the ML models could accurately predict them. The study could improve by modifying the models to handle a higher number of loans.

7 Conclusions

In this research paper, we present an FL prototype for the credit risk assessment of mortgages. We evaluate this prototype with an empirical dataset and a scenario analysis consisting of five scenarios. We find that smaller financial institutions could benefit significantly from collaboration with others through FL. On average, our FL prototype improved accuracy, recall, precision, and F1 scores by 4.02, 1.84, 7.93, and 5.59 percentage points respectively.

The work presented in this paper contributes to the existing literature on the use of FL in financial services. In particular, our study contributes the following:

1. We present a prototype that uses FL for credit risk assessment of mortgages;
2. We demonstrate empirically the potential benefit of using FL for the credit risk assessment of mortgages.

Acknowledgements

We would like to thank to Daniel Howard Maclennan and Mohammad Ansarin for their valuable feedback.

This work has been supported by the European Union (EU) within its Horizon 2020 programme, project MDOT (Medical Device Obligations Taskforce), Grant agreement 814654, from the Kopernikus-project "SynErgie" by the German Federal Ministry of Education and Research (BMBF), from PayPal and the Luxembourg National Research Fund FNR (P17/IS/13342933/PayPal-FNR/Chair in DFS/Gilbert Fridgen), as well as from the Luxembourg National Research Fund (FNR) – FiReSpARX Project, ref. 14783405. The experiments presented in this paper were carried out using the HPC facilities of the University of Luxembourg (Varrette et al., 2014)– see hpc.uni.lu

References

Altman, E. I. (2002). "Managing Credit Risk: A Challenge for the New Millennium". In: *Economic Notes* 31.2, pp. 201–214. DOI: 10.1111/1468-0300.00084. eprint: https://onli nelibrary.wiley.com/doi/pdf/10.1111/1468-0300.00084. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0300.00084.

Aïvodji, U. M., S. Gambs, and A. Martin (2019). "IOTFLA : A Secured and Privacy-Preserving Smart Home Architecture Implementing Federated Learning". In: *2019 IEEE Security and Privacy Workshops (SPW)*, pp. 175–180. DOI: 10.1109/SPW.2019.00041.

Bank, E. C. (2010). *Memorandum of understanding on the exchange of information among national central credit registers for the purpose of passing it on to reporting institutions*. Tech. rep.

Banking Supervision, B. C. on (2018). *Pillar 3 disclosure requirement - updated framework*. Tech. rep.

Bansal, A., R. J. Kauffman, and R. R. Weitz (1993). "Comparing the Modeling Performance of Regression and Neural Networks as Data Quality Varies: A Business Value Approach". In: *Journal of Management Information Systems* 10.1, pp. 11–32. ISSN: 07421222. URL: http://www.jstor.org/stable/40398029.

Borgman, C. L. (2012). "The conundrum of sharing research data". In: *Journal of the American Society for Information Science and Technology* 63.6, pp. 1059–1078. DOI: 10.1002/asi.22634. eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.22634. URL: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.22634.

Chen, N., B. Ribeiro, and A. Chen (2016). "Financial credit risk assessment: a recent review". In: *Artificial Intelligence Review* 45. DOI: 10.1007/s10462-015-9434-x.

Chollet, F. et al. (2015). *Keras*. https://keras.io.

Ekbia, H., M. Mattioli, I. Kouper, G. Arave, A. Ghazi, T. Bowman, V. Suri, A. Tsou, S. Weingart, and C. Sugimoto (2015). "Big Data, Bigger Dilemmas: A Critical Review". In: *Journal of the Association for Information Science and Technology* 66. DOI: 10.1002/asi.23294.

Federal Reserve (2021). *Federal Reserve Economic Data (FRED)*. Accessed: 2021-08-11. URL: https://fred.stlouisfed.org/.

Freddie Mac (2021a). *Freddie Mac's House Price Index (FMHPI)*. Accessed: 2021-08-11. URL: https://www.freddiemac.com/research/indices/house-price-index.

Freddie Mac (2021b). *Single Family Loan-Level Dataset*. Accessed: 2021-08-11. URL: http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.page.

Galindo, J. and P. Tamayo (2000). “Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications”. In: *Computational Economics* 15, pp. 107–43. DOI: 10.1023/A:1008699112516.

Heitfield, E. (2009). “Parameter Uncertainty and the Credit Risk of Collateralized Debt Obligations”. In: *Risk Management*.

Hochreiter, S. and J. Schmidhuber (1997). “Long short-term memory”. In: *Neural computation* 9.8, pp. 1735–1780.

Israël, J.-M., V. Damia, R. Bonci, and G. Watfe (2017). *The Analytical Credit Dataset - A magnifying glass for analysing credit in the euro area*. Occasional Paper Series 187. European Central Bank. URL: https://ideas.repec.org/p/ecb/ecbops/2017187.html.

Kaissis, G., A. Ziller, J. Passerat-Palmbach, T. Ryffel, D. Usynin, A. Trask, I. Lima, J. Mancuso, F. Jungmann, M.-M. Steinborn, A. Saleh, M. Makowski, D. Rueckert, and R. Braren (2021). “End-to-end privacy preserving deep learning on multi-institutional medical imaging”. In: *Nature Machine Intelligence*, pp. 1–12. DOI: 10.1038/s42256-021-00337-8.

Kearns, G. S. and A. L. Lederer (2004). “The impact of industry contextual factors on IT focus and the use of IT for competitive advantage”. In: *Information & Management* 41.7, pp. 899–919. ISSN: 0378-7206. DOI: https://doi.org/10.1016/j.im.2003.08.018. URL: https://www.sciencedirect.com/science/article/pii/S0378720603001459.

Kroese, D. P., T. J. Brereton, T. Taimre, and Z. I. Botev (2014). “Why the Monte Carlo method is so important today”. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 6.

McMahan, H. B., E. Moore, D. Ramage, and B. A. y Arcas (2016). “Federated Learning of Deep Networks using Model Averaging”. In: *CoRR* abs/1602.05629. arXiv: 1602.05629. URL: http://arxiv.org/abs/1602.05629.

McMahan, H. B., E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas (2017). “Communication-Efficient Learning of Deep Networks from Decentralized Data”. In: *Artificial Intelligence and Statistics*, pp. 1273–1282.

Murphy, A. (2008). “An analysis of the financial crisis of 2008: causes and solutions”. In: *An Analysis of the Financial Crisis of*.

Redman, T. C. (1995). “Improve Data Quality for Competitive Advantage”. English. In: *Sloan management review* 36.2. Copyright - Copyright Sloan Management Review Association, Alfred P. Sloan School of Management Winter 1995; Last updated - 2021-

09-09; SubjectsTermNotLitGenreText - United States–US, p. 99. URL: https://www.proquest.com/scholarly-journals/improve-data-quality-competitive-advantage/docview/224971040/se-2?accountid=41819.

Robbins, H. and S. Monro (1951). "A stochastic approximation method". In: *The annals of mathematical statistics*, pp. 400–407.

Rosenblatt, M. (1956). "Remarks on Some Nonparametric Estimates of a Density Function". In: *The Annals of Mathematical Statistics* 27.3, pp. 832 –837. DOI: 10.1214/aoms/1177728190. URL: https://doi.org/10.1214/aoms/1177728190.

Saputra, Y. M., D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, M. D. Mueck, and S. Srikanteswara (2019). "Energy Demand Prediction with Federated Learning for Electric Vehicle Networks". In: *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6. DOI: 10.1109/GLOBECOM38437.2019.9013587.

Sezer, O. B., M. U. Gudelek, and A. M. Ozbayoglu (2020). "Financial time series forecasting with deep learning : A systematic literature review: 2005–2019". In: *Applied Soft Computing* 90, p. 106181. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2020.106181. URL: https://www.sciencedirect.com/science/article/pii/S1568494620301216.

Shingi, G. (2020). "A federated learning based approach for loan defaults prediction". In: *2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 362–368. DOI: 10.1109/ICDMW51313.2020.00057.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56, pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

Taleb, N. N. (2020). "Statistical consequences of fat tails: Real world preasymptotics, epistemology, and applications". In: *arXiv preprint arXiv:2001.10488*.

TensorFlow Federated (2018). Version 0.20.0. URL: https://github.com/tensorflow/federated.

United States Bureau of Labor Statistics (2021). *United States Bureau of Labor Statistics' Local Area Unemployment Statistics (LAUS)*. Accessed: 2021-08-11. URL: https://www.bls.gov/lau/.

Varrette, S., P. Bouvry, H. Cartiaux, and F. Georgatos (2014). "Management of an Academic HPC Cluster: The UL Experience". In: *Proc. of the 2014 Intl. Conf. on High Performance Computing & Simulation (HPCS 2014)*. Bologna, Italy: IEEE, pp. 959–967.

Walczak, S. (2001). "An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks". In: *Journal of Management Information Systems* 17.4, pp. 203–222. DOI: 10.1080/07421222.2001.11045659. eprint: https://doi.org/10.1080/07421222.2001.11045659. URL: https://doi.org/10.1080/07421222.2001.11045659.

Yang, W., Y. Zhang, K. Ye, L. Li, and C. Xu (2019). "FFD: A Federated Learning Based Method for Credit Card Fraud Detection". In: *BigData Congress*.

Zheng, W., L. Yan, C. Gou, and F.-Y. Wang (2020). "Federated Meta-Learning for Fraudulent Credit Card Detection". In: *IJCAI*.

Zuiderwijk, A., M. Janssen, K. Poulis, and G. van de Kaa (2015). "Open Data for Competitive Advantage: Insights from Open Data Use by Companies". In: *Proceedings of the 16th Annual International Conference on Digital Government Research*. dg.o '15. Phoenix, Arizona: Association for Computing Machinery, 79–88. ISBN: 9781450336000. DOI: 10.1145/2757401.2757411. URL: https://doi.org/10.1145/2757401.2757411.

Research Paper 4 – *Federated Learning as a Solution for Problems Related to Intergovernmental Data Sharing*

Full Title:

Federated Learning as a Solution for Problems Related to Intergovernmental Data Sharing

Authors:

Kilian Sprenkamp, Joaquín Delgado Fernández, Sven Eckhardt, Liudmila Zavolokina

Published in: Hawaii International Conference in System Sciences (HICSS) (2023)

Keywords:

Federated Learning, Artificial Intelligence, eGovernment, Data sharing challenges

DOI:

https://hdl.handle.net/10125/102838

Abstract

To address global problems, intergovernmental collaboration is needed. Modern solutions to these problems often include data-driven methods like artificial intelligence (AI), which require large amounts of data to perform well. However, data sharing between governments is limited. A possible solution is federated learning (FL), a decentralised AI method created to utilise personal information on edge devices. Instead of sharing data, governments can build their own models and just share the model parameters with a centralised server aggregating all parameters, resulting in a superior overall model. By conducting a structured literature review, we show how major intergovernmental data sharing challenges like disincentives, legal and ethical issues as well as technical constraints can be solved through FL. Enhanced AI while maintaining privacy through FL thus allows governments to collaboratively address global problems, which will positively impact governments and citizens.

1 Introduction

Even though there are approaches to allying with other countries, objectively, sovereign nation-states exercise power over a population of citizens within their territorial borders. With the increasing impact of digital technology and the rise of the internet as a "borderless space", the role of traditional borders in the digital realm is questioned more and more often. Although the very essence of the internet is to connect users and devices beyond borders, countries attempt to preserve their sovereignty by subjecting cyberspace to their own national rules.

An area where sovereignty is widely pursued among different countries is data sharing. While sharing data allows enhanced analytics and value generation, the 2022 World Economic Forum underlined the problem that data sharing is impractical as data is stored in different legacy system silos (Antonio, 2022). Further, legitimate reasons making data sharing complicated are disincentives due to the collective action theory (Olson, 1965) and data sharing being unethical (Ward and Sipior, 2010), especially when personal information is involved. In addition, legal uncertainties exist when data is shared between nations created through legislation like the General Data Protection Regulation (2018).

Still, technological advances do not stop, and governments need to keep up with the waves of innovation for economic, social and political reasons. While industry 4.0 is a very common term for the technological advancement in industry, eGovernment 3.0 (eGOV3.0) is the term used to describe the ever-increasing use of disruptive information and communication technology (ICT), such as artificial intelligence (AI) in governments. The term eGovernment 1.0 describes the use of ICT for the realisation of public services (Lachana et al., 2018). In addition, eGovernment 2.0, focuses on the ICT-enabled participation of citizens, while eGOV3.0 uses more advanced and data-driven technologies to solve societal problems through collected data (Lachana et al., 2018).

In order to fulfil the aspiration of AI to solve societal problems vast amounts of data are needed (Duan et al., 2019), which can be acquired through intergovernmental data sharing. From a global perspective, this need for data immediately creates tension between governments' interests and incentives, i.e., one government might want to pursue open data sharing, and another might seek to maintain their data private. This tension is illustrated by the statement of Germany's former health minister regarding the World Health Organisation's (WHO's) potential to place sanctions on countries that do not share their data during disease outbreaks (Wheaton and Martuscelli, 2021). While such a pandemic treaty exists, it is only actively supported by 25 countries (WHO, 2021), making the creation of accurate AI models in regard to Epidemiology difficult. Therefore, the question arises of how to access data without sharing it in order to collaboratively solve global problems. It can be argued that we can still build AI methods on private data. However, to advance the newly created eGOV3.0, we need to ensure that AI models operate as well as possible, as the performance of these systems has a far-reaching influence on governments and directly on citizens' lives.

One recent approach that promises to solve these problems is federated learning (FL) (McMahan et al., 2017). The core idea of FL is that individual entities build their own AI models and share them at a centralised point. Another AI model is built that aggregates the individual models. At no point in time is the data of any individual entity shared. In general, FL shows that the *federated model* performs better than *individual models*, which are built solely on the data of a single entity. However, federated models generally perform worse than the *oracle model*, a model built with all available data stored in a single silo. Nonetheless, it is often impractical to build an oracle model due to the limitations of data sharing (Jordan and Mitchell, 2015), leaving open the question of which technological approach would be the most fitting.

In our paper, we propose to use FL to enable better intergovernmental collaborations. We, therefore, investigate the research gap regarding how FL can be used for international eGOV3.0 in use cases where data cannot be shared. The following research question (RQ) is formulated:

1. *How can federated learning address the problem of data sharing in intergovernmental collaboration?*

To answer this RQ, we investigate how challenges in data sharing listed by the Organisation for Economic Co-operation and Development (OECD) in OECD (2019) can be mitigated through FL. We choose the challenges named by OECD (2019) as a scientific framework listing intergovernmental data sharing challenges in the context of AI does not exist to our knowledge. We analyse these challenges through a structured literature review, aiming to propose FL as a solution to the challenges of intergovernmental data sharing.

The paper is structured as follows. In the subsequent chapter, we present related work and the background to the study. In Chapter 3, we explain our methodology. We present our results in Chapter 4. In Chapter 5, we discuss the implications of our results. We then conclude the study with an outlook in Chapter 6.

2 Related Work and Background

2.1 Intergovernmental Data Sharing

Data sharing is an ever-increasing factor for intergovernmental collaboration and success (Wiseman, 2020). Examples of successfully created AI applications trained on intergovernmental data include health, mobility and the social sector (Wiseman, 2020). Yet, there are legitimate national, public and private interests (OECD, 2019), making data sharing between administrations disincentive, unethical, legally uncertain or impractical due to technology constraints. We focus specifically on these four issues as they are well researched within the scientific literature. Thus, OECD (2019) functions as an extension focusing on practical intergovernmental challenges.

First, for intergovernmental collaboration, it is of great importance to take into account governments' counterincentives to share data, as sharing data might conflict with other policy goals (OECD, 2019). This can be due to information asymmetry that arises between information-poor and information-rich countries and can result in negative consequences for each type of country (Clarkson et al., 2007). Information-poor countries are often in a weaker position to negotiate data sharing agreements (Clarkson et al., 2007) and are thus inclined to make less advantageous concessions. In contrast, information-rich countries might have a counterincetive to share their data as they want to maintain their strong economic position. This reluctance to share data emerges from what is known as the "free rider" problem, where data is a non-exclusive public good and information-rich countries have to accept the risk of information-poor countries utilising their good free of charge (OECD, 2019). Due to "free riding" on the goods provided between organisation the allocation of public goods becomes ineffective, which is known as the collective action theory (Olson, 1965). Collective action thus results in difficulties for inter-organisational cooperation. While Olson (1965) focuses on inter organisational cooperation the theory has been expanded to problems regarding intergovernmental cooperation, e.g., Aspinwall and Greenwood (2013) for cooperation within the European Union (EU) allowing "free riding" of public goods provided by sovereign nation-states.

Second, some data has special privacy rights, such as personal information. Sharing this data can create ethical concerns. Data breaches from the private sector, like Facebook or Meta (Isaak and Hanna, 2018), and the public sector, like the disclosure of the records of 191 million voters in the United States (Bennett, 2016), decreases user trust, and data subjects are less likely to share data again (Pingitore et al., 2017). Therefore, countries need to ensure transparency, disclosure, control and notification in case of the maltreatment of citizens' data (Isaak and Hanna, 2018).

Third, the fear of legal consequences in a fragmented regulatory landscape limits the ability of data sharing. This is amplified by legal uncertainties over who controls the data and under which legislation and jurisdiction it falls (Ward and Sipior, 2010). Especially complicated is the distribution of data among multiple administrations with conflicting bilateral agreements. An example could be the cross-border transfer of data between EU countries, Japan and the United States, which is not currently possible. Right now, the EU only recognises nine non-EU countries as providing adequate protection for saving data. Japan is among them, but the United States is not (General Data Protection Regulation, 2018). Moreover, the EU can issue fines to any organisation not complying with General Data Protection Regulation (2018), creating a further monetary disincentive to share data based on EU law binding to nations inside and outside the EU.

Last, while data should be distributed in according to the FAIR (findability, accessibility, interoperability, reusability) principle (Wilkinson et al., 2016) several technological challenges and threats for governments can occur in data sharing. While in the age of cloud computing the cost of storing, copying and analysing data has shrunk, open data provision still involves significant costs for collecting, preparing, sharing, scaling, maintaining and updating data (Chen and Zhang, 2014; Johnson, 2016). Another challenge is the varying quality of data due to inconsistency and incompleteness and the resulting need for standardisation when data is stored in multiple locations (Chen and Zhang, 2014; Mikhaylov et al., 2018). Data sharing further creates multiple entry points into a system, decreasing data security (Chen and Zhang, 2014; Chen and Zhao, 2012).

These challenges, arising from national, public and private interests, are especially relevant as they hinder the sharing of data and thus the creation of AI trained on data recorded in multiple countries. This, in turn, impacts the advance of eGOV3.0 and the realisation of its benefits to society as a whole.

2.2 Federated Learning

A possible solution to the challenges of data sharing for the creation of AI models was proposed by McMahan et al. (2017), who, while leveraging data utility, maintained a clear separation between data owners. This solution, namely FL, relies on the distribution of models across different databases instead of the classical machine learning (ML) example, where all the data is stored in a single silo. By distributing the models, the authors separated the model and the data, keeping the latter isolated and at the data owners' selected location without revealing it. Since then, applications in health (Xu et al., 2021), banking (Yang et al., 2019) or smart cities (Jiang et al., 2020) have been subject to research. FL thus provides incentives in terms of privacy, security, legal and economic benefits for users (Yang et al., 2019). To the extent of our knowledge, one framework on how to apply FL in eGOV3.0 use cases (Guberović et al., 2022) has been created, which includes the specification of client, server, model and application programming interface requirements at the start of a project. Further, the accountability of FL in government to overcome legislative constraints has been researched, pointing out engineering requirements, i.e., architecture design and management requirements, i.e., trust among actors (Balta et al., 2021).

Originally, McMahan et al. (2017) proposed FL while working at Google, utilising decentralised data stored on multiple edge devices for tasks like image or voice recognition. Now, the original technique is also termed "horizontal FL" as the data of each client shares the same feature space. We visualise FL incorporated into eGOV3.0 in cross-silo (Wiseman, 2020) use cases (see Figure 1). The term cross-silo refers to different data storages on a national or sub national level, which due to the challenges of intergovernmental data sharing, explained in Section 2.1 cannot be exchanged between nations.

The algorithm is initialised with a base model sent to all clients (i.e., nation-states) by the server (i.e., an intergovernmental organisation or intergovernmental collaboration project). This step is not part of the repeatedly performed steps and is therefore named step 0. In step 1, each client starts the training process from the base model using their own data. In step 2, the difference between the base model parameters and the client model parameters is sent to the server, but each client's data is not shared. While sharing model weights is also a form of information sharing, security techniques like *secure*

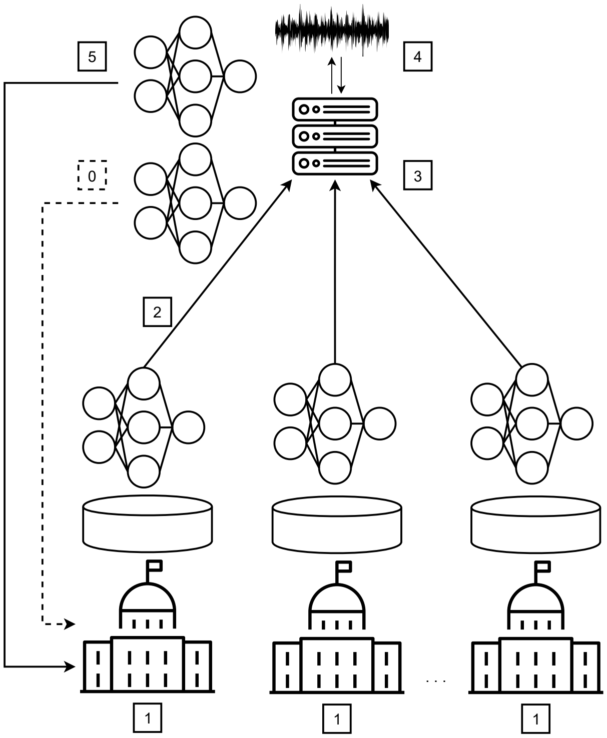


Figure 1: Federated Learning in eGOV3.0

aggregation and *differential privacy* keep data secure and private. The communication channel between client and server is ciphered via *secure aggregation* (Bonawitz et al., 2017) and thus made secure. *Secure aggegration* thus allows for the transfer of model parameters between parties which do not trust each other, other clients or the server are not capable of obtaining the model weights other clients send through the federated system. During step 3, the server utilises *federated averaging* to calculate a weighted mean of all differences. The weight of each client is determined by the amount of data used to train the model. Then, in step 4, the server adds random noise to the aggregated model. By adding random noise, privacy is ensured, meaning that the prior steps cannot be reverse-engineered. This procedure is known as *differential privacy* (Agarwal et al., 2018) and is also used in standard ML algorithms. Differentially private models are defined by the tuple (ϵ,δ), where ϵ defines the impact of each individual piece of information on the results of the analysis. In other words, low ϵ indicates a robust system where the outcome is not represented by any particular client data. Additionally, δ regulates the likelihood of a data breach occurring. Step 4 is optional, but if chosen, the model becomes a secure federated model, allowing governments to keep their data private. Finally, in step 5, the private aggregated model is sent to the clients. Steps 1 to 5 represent one round of federated training, which is repeated until the federated model converges.

For the paper we focus how FL can solve problems related to external intergovernmental data sharing meaning collaboration happening between multiple sovereign nation-states. However, partly we consider that individuals, as well as companies, give their data to the respective nation-state they are located in, e.g., multiple states creating a model for the prediction of the effect of climate change based on CO_2 data collected from companies or households.

While FL has been proposed by McMahan et al. (2017) to solve the problem of data sharing on edge devices, there have been limited attempts to adapt this technique to the public sector. It is especially unclear how the specific challenges of data sharing among governments can be solved.

3 Method

We conducted a structured literature review to investigate how FL can solve data sharing problems identified by OECD (2019). We utilised the challenges provided by the OECD as a single scientific framework for intergovernmental data sharing does not exist to our knowledge. However, the OECD study results from workshops attended by data professionals and policy leaders from various countries and industries, ensuring a diverse and holistic view. The challenges of data sharing identified by OECD (2019) are grouped into three categories and various subcategories (see table 1) and formulated with regards to the growing importance of AI. During the literature search process, we focused on information systems and computer science literature, and we applied forward as well as backward searches (Webster and Watson, 2002). We utilised the search string `{"Federated Learning" AND "Data" AND "Challenge*"}` in the disciplines of Information Systems, Computer, Decision, Manangement and Social Science, yielding a total of 879 results. We searched through the scopus, IEEE, ACM and the AIS library database. As pre-selection criteria, we analysed if the given article discusses the challenges named by OECD (2019) based on the abstract, 43 articles were thus chosen for further analysis. We mapped 14 articles to the challenges of the OECD (2019), a forward and backward found further 7 articles resulting in 21 articles selected.

4 Results

We identified solutions to the sub-challenges (OECD, 2019) by conducting a structured literature review. We found solutions for eight of the 12 sub-challenges defined, while two of the sub-challenges were partially solved and two of the sub-challenges remained unsolved (see Table 1).

[1] Balancing the benefits of data openness with legitimate interests, policy objectives and risks.

With FL, the security and confidentiality breaches in data sharing [1.1] can be avoided, as FL alleviates the need to share data. FL does so through the range of security and privacy techniques (Mothukuri et al., 2021). We like to point out two core approaches *differential privacy* (Agarwal et al., 2018) and *secure aggregation* (Bonawitz et al., 2017). While *differential privacy* secures the system from being reverse-engineered, *secure aggregation* secures the system by ciphering the communication channel between the clients and the server. Moreover, concerning the importance of personal information, hierarchical FL settings (Abad et al., 2020) enable FL to be applied on multiple levels. A country could thus give citizens or companies control over their data while still profiting from AI being trained by international projects.

The violation of privacy and property rights [1.2] is, according to OECD (2019), based on contractual agreements. The violation of these agreements can lead to fines. Moreover, sharing data prematurely can reduce the chance of creating intellectual property. FL offers a technological pathway for entities to comply with these contractual agreements (Li et al., 2020a; Li et al., 2021), thus preventing them from violating contractual clauses and being exposed to ensuing fines or the premature revelation of intellectual property.

Regarding mitigating the difficulty of applying risk management approaches [1.3], we currently see limited potential in FL to solve this challenge.

Table 1: FL solutions to data sharing challenges (OECD, 2019)

Challenge	Sub Challenge	Solved by FL	Proposed Solution
[1] Balancing the benefits of data openness with legitimate interests, policy objectives and risks	[1.1] Security risks and confidentiality breaches	solved	• Security and privacy techniques (Agarwal et al., 2018; Bonawitz et al., 2017; Mothukuri et al., 2021) • Hierarchical federated learning (Abad et al., 2020)
	[1.2] Violation of privacy and intellectual property	solved	• Workaround for contractual agreements (Li et al., 2020a; Li et al., 2021)
	[1.3] Difficulty of risk management approaches	not solved	• N/A
	[1.4] Cross-border data access and sharing	solved	• Unnecessity of cross-border data sharing (Truong et al., 2021; Yang et al., 2021; Yang et al., 2019)
[2] Trust and empowerment for the effective re-use of data across society	[2.1] Supporting and engaging communities	not solved	• N/A
	[2.2] Fostering data-related infrastructures and skill	partially solved	• Communication cost (Li et al., 2020b; McMahan et al., 2017) • FL toolkits (Ziller et al., 2021)
	[2.3] Lack of common standards for data sharing and re-use	solved	• System heterogeneity (Mitra et al., 2021) • Vertical federated learning (Yang et al., 2019) • Federated transfer Learning (Chen et al., 2020)
	[2.4] Data quality	solved	• FL on noisy data (Passerat-Palmbach et al., 2020; Tuor et al., 2021)
[3] Misaligned incentives, and limitations of current business models and markets	[3.1] Externalities of data sharing, re-use and misaligned incentives	solved	• Incentives of FL (Kang et al., 2019; Yu et al., 2020)
	[3.2] Limitations of current business models and data markets	solved	• FL as a business model (Balta et al., 2021; Manoj et al., 2022; Yang et al., 2019)
	[3.3] The risks of mandatory access to data	partially solved	• Evaluation of samples (Ziller et al., 2021)
	[3.4] Uncertainties about data ownership	solved	• Data ownership is explicit (Liu et al., 2020; Shae and Tsai, 2018) • Model ownership is explicit on technical level (Liu et al., 2021)

From a legal perspective, cross-border data sharing [1.4] can be complicated due to regulations like the General Data Protection Regulation (2018). FL allows the training of AI without the need to share data across borders. Yang et al. (2019) proposed the training of federated models between Chinese and American companies. Similarly, Yang et al. (2021) presented an FL system using data from China, Japan and Italy to predict SARS-

CoV-2 from chest computed tomography images. However, Truong et al. (2021) point out that due to to the exchange of model weights and the resulting threat of backward engineering, FL, without any security and privacy techniques is not conform with General Data Protection Regulation (2018) in Europe. Therefore, FL can only be utilised with privacy and security preserving techniques when utilising data from multiple countries.

[2] Trust and empowerment for the effective re-use of data across society

FL cannot help to create more engagement in open data communities [2.1] as it reduces the need to share data. However, we can see that the shared model training creates a community aspect. To our knowledge, this has not yet been researched.

We found that FL cannot solve problems related to data infrastructure or skills [2.2]. Generally, the technique requires a more complicated setup (Li et al., 2020b). FL profits from dividing the computational cost across multiple clients. Yet, the cost of communication is high in FL, as clients need to communicate continuously. This cost is practically non-existent in normal ML (McMahan et al., 2017). The reduction of costs associated with FL is currently under research, and several solutions have been proposed (Li et al., 2020b). Further, due to the increased complexity and novelty of FL, tool kits like `TensorFlow Federated` based on McMahan et al. (2017) and `PySyft` (Ziller et al., 2021) have not seen wide adoption compared to other ML frameworks. Nonetheless, we see possibilities for less skilled countries to profit as they prefer to use federated models rather than training models themselves.

The lack of data standardisation [2.3] reflects two core challenges of FL: system and statistical heterogeneity. Solutions to this issue exist already. System heterogeneity is described as different hardware being used among clients, leading to a slower training process. For example, Mitra et al. (2021) propose re-using parts of the model during the training process, such as gradients of the learned network or the specification of concrete learning rates for individual clients depending on the hardware. Statistical heterogeneity refers to different data features being stored or features having non-identical distributions across clients. In this case, vertical FL (Yang et al., 2019) can be used, which allows for the training of models with different feature spaces. Moreover, it is possible to apply transfer FL, meaning that a model is retrained for a different learning task, benefiting from the knowledge of the previously learned task. Chen et al. (2020) employ this technique by first training a model for activity recognition for smartwatches, which is then transferred to the task of predicting Parkinson's disease.

The OECD notes that poor data quality [2.4] will lead to poor analytics. While FL cannot improve the data quality of clients, entities with poor data quality can profit from the federated model. Examples can be found within the medical field of genomics or mental health, where large amounts of noisy data can be found (Passerat-Palmbach et al., 2020). In this case clients with poor data can profit from the federated model and the contribution made by other clients with better quality data. Still, the clients with poor data quality will decrease the overall model quality. A solution to this challenge is proposed by Tuor et al. (2021), each client evaluates their data set with a benchmark model trained on high quality data. For bad quality data the model will be incapable of making a prediction, generating a high loss value. These data points will not be further utilised for training.

[3] Misaligned incentives, and limitations of current business models and markets

A central problem within data sharing is misaligned incentives [3.1] between information-rich and information-poor countries. The problem of incentivising information-rich clients to participate in FL has been well researched (Kang et al., 2019; Yu et al., 2020). These methods typically offer a reward for participating within the federated system. Hence, an entity could earn depending on how much value was brought to the federated model.

Implementing FL could significantly reduce the need for data markets [3.2], as data can be kept by the owner. Moreover, according to the OECD, the ex-ante evaluation of the economic potential of data is challenging. However, given the previously shown incentive schemes of FL, it is possible to track participation in an FL project. Yang et al. (2019) estimate that FL will evolve into a business model where participants in an FL project can profit from the value they contribute to the model. FL thus allows participants to pursue joint business activities (Balta et al., 2021). An example of such a joint business activity is given by Manoj et al. (2022) training a model for predicting the yield of crop. This model can be utilised by multiple stakeholders, e.g., farmers for revenue estimates, banks and insurances for mitigating risks and governments for setting export prices.

In a federated setting, mandatory data access [3.3] can be kept to a minimum. For example, the `PySyft` package (Ziller et al., 2021) within `Python` allows for viewing a limited number of samples of each client's data to optimise the federated model. Accessing all available data points is, in theory, not necessary and, due to the number of data points available, not always feasible while training AI models.

Finally, the OECD notes a loss of data ownership [3.4] as an emerging challenge of sharing data. With FL, the ownership of a data point remains unaffected as it is not shared across multiple sources. For example, Shae and Tsai (2018) propose storing medical information on a blockchain for training federated models. Thus, the ownership of data cannot be falsified. In a similar manner, Liu et al. (2020) proposed a traffic flow prediction model utilising data from government organisations, smart devices, private persons as well as private companies like Uber or Didi. For each of these entities, the data ownership is unambiguous. Moreover, it is possible to verify the ownership over a trained federated model by implementing a watermarking technique, thus, the contribution and resulting ownership of clients is recorded through the watermark (Liu et al., 2021). However, the watermarking technique solely clarifies ownership on a technical and not legal level. To our knowledge the legal ownership of federated models remains unsolved.

5 Discussion

This study aimed to address the research gap regarding how FL can be used in international eGOV3.0 use cases where data sharing is complicated or unfeasible. Previous research has shown that data sharing is limited due to being disincentive (Olson, 1965), unethical (Isaak and Hanna, 2018; Pingitore et al., 2017), legally uncertain (Ward and Sipior, 2010) or impractical due to technical challenges (Chen and Zhang, 2014; Chen and Zhao, 2012; Johnson, 2016; Mikhaylov et al., 2018). Since McMahan et al. (2017) first proposed FL, a vast number of publications have appeared in the field, including applications in health, banking and smart cities. However, research in the area of eGOV3.0 is limited. While Guberović et al. (2022) created a framework specifying component requirements for government FL projects and Balta et al. (2021) analysed the accountability of FL in government, we analysed how FL can solve the problem of intergovernmental data sharing. We did so by conducting a structured literature review, which served the purpose of analysing how FL can solve challenges identified by OECD (2019). In doing so, we were able to answer the given RQ: *How can federated learning address the problem of data sharing in intergovernmental collaboration?*

First, FL can help to deal with legal (Ward and Sipior, 2010) and ethical (Isaak and Hanna, 2018; Pingitore et al., 2017) issues around data sharing. We provided evidence regarding how the sub-challenges [1.1], [1.2], [1.4] and [3.4] identified by the OECD can be solved. FL significantly reduces the need for data sharing agreements to build AI (Li et al., 2020a; Li et al., 2021), which also applies to cross-border data sharing (Truong et al., 2021; Yang et al., 2021; Yang et al., 2019). A further consequence is that data ownership cannot be falsified, as the data is stored at the owners' selected location. Moreover, FL, mainly through *differential privacy* (Agarwal et al., 2018) and *secure aggregation* (Bonawitz et al., 2017), allows for secure model training and keeping data private. FL, thus provides a trusted technology that is ideal for intergovernmental use cases.

Second, we showed how technological constraints in data sharing (Chen and Zhang, 2014; Chen and Zhao, 2012; Johnson, 2016; Mikhaylov et al., 2018) can be overcome using FL, especially issues regarding the sub-challenges of standardisation of data [2.3] and data quality [2.4]. Mitra et al. (2021) show methods that allow federated models to be trained in system heterogeneous settings. This is beneficial for FL in intergovernmental settings as countries, companies and citizens can partake in FL projects without the need for special hardware. Vertical FL (Yang et al., 2019) and transfer FL (Chen et al., 2020) allow the training of AI models on non-standardised data sets, and they can even leverage data that was not recorded for the task they have been employed to solve. Intergovernmental collaboration can thus profit from data stored by all types of entities and further re-use data effectively. Additionally, data from both information-poor and information-rich countries can be utilised to contribute to FL projects. Information-poor countries are not limited to their own data sources anymore and can contribute and profit from the federated model. Still, while federated learning shows potential and has been implemented on a scientific basis till now applications of FL in an intergovernmental context in real-world scenarios are not known to us. This could be the case due to technological challenges solely being solved in the literature but not in real-world scenarios.

Last, FL can evolve into a business model (Balta et al., 2021; Yang et al., 2019), which gives entities an incentive to take part in intergovernmental projects. This mitigates disincentives in data sharing caused by “free riding” and the problem of inefficiency due to collective action (Olson, 1965). Consequently, we demonstrated how sub-challenges [3.1] and [3.2] could be solved through FL. With incentive mechanisms developed for FL (Kang et al., 2019; Yu et al., 2020), both information-rich and information-poor entities can be motivated to participate in an FL project by offering a reward. For example, an intergovernmental climate model could be possible where different stakeholders, e.g. countries, companies or single individuals could earn money or CO_2 credits based on the revenue or value that an intergovernmental FL project generates. Still, large contributors would earn the most, but also, less funded entities can profit fairly. We consider the incentives to partake in FL to be superior to the incentives for partaking in data sharing. When using FL, the ownership of the data [3.4] remains intact for each FL project an entity might participate in. In contrast, for standard data sharing, as soon as data is distributed, it becomes unclear who the real owner is. Therefore, “free riding” as described in the collective action theory (Olson, 1965) and the resulting inefficiency can be mitigated on a technical level in theory.

However, not all problems related to data sharing can be solved through FL. Sub-challenges [2.2] and [3.3] are only partially solved, while sub-challenges [1.3] and [2.1] are not solved. There are two key problems. First, implementing FL infrastructure is more complicated, with higher communication costs. Even with a larger adoption, this will cause challenges in eGOV3.0. Second, although ownership of FL is actively researched from a technological point of view (Liu et al., 2021), we see problems on the organisational level, in which the geolocation and the controlling entity of the server aggregating the model will play a central role. The entity controlling the federated model could cut off countries in an intergovernmental collaboration project without a democratic process. Especially from a political realism perspective, it is unlikely that nations which do not trust each other will participate in a joint FL project. Future research could consider governments’ willingness to participate in projects that benefit various multinational stakeholders. Within this analysis, the difference between incentives of information-rich and information-poor countries is likely to play a key role.

6 Conclusion

This study conducts a structured literature review to show how FL can function as a solution to various challenges related to intergovernmental data sharing. FL enables the training of models in a decentralised manner and can thus reduce incentive, legal, ethical and practical challenges in intergovernmental data sharing. Nevertheless, secondary problems of a technical and organisational nature arise.

The contribution of this study is threefold. First, we present a state-of-the-art AI method to overcome the problem of intergovernmental data sharing. This serves as a basis for FL research in international eGOV3.0, which we hope will influence both governments and citizens. Second, we contribute to the existing literature on FL, providing a structured review on how FL should be utilised in eGOV3.0, focusing on the aspect of data sharing. We hope this will enhance the research output of real-life use cases in and outside the eGOV3.0 space. Third, we show a new possibility of how to mitigate inefficiency created by the collective action theory (Olson, 1965).

Moreover, our study has the limitation of solely focusing on the challenges of data sharing provided by the OECD (2019). We estimate that challenges described by other authorities will be similar, but adding challenges from authorities of different cultural or economic origins would create an even more holistic and diverse picture.

Considering the opportunities and challenges highlighted in this study, we find substantial potential for the information systems and related research communities. We suggest creating federated systems on proprietary, potentially unbalanced data from multi-governmental stakeholders. Dedicated qualitative research could be done, drawing insights from workshops or stakeholder interviews to further investigate the potential of FL in eGOV3.0. Moreover, at the organisational level, it is necessary to determine the owner of the server that creates the aggregated model. Finally, the higher infrastructure costs and skill levels of users in FL need to be considered in further research.

References

Abad, M. S. H., E. Ozfatura, D. Gunduz, and O. Ercetin (2020). “Hierarchical federated learning across heterogeneous cellular networks”. In: *ICASSP*. IEEE.

Agarwal, N., A. T. Suresh, F. X. X. Yu, S. Kumar, and B. McMahan (2018). "cpSGD: Communication-efficient and differentially-private distributed SGD". In: *Advances in Neural Information Processing Systems*.

Antonio, N. (2022). "The public sector must accelerate digital transformation – or risk losing sovereignty and trust". In: *Politico*. URL: https://www.weforum.org/agenda/2022/05/the-public-sector-must-accelerate-digital-transformation-or-risk-losing-sovereignty-and-trust/ (visited on 05/23/2022).

Aspinwall, M. and J. Greenwood (2013). *Collective action in the European Union: interests and the new politics of associability*. Routledge.

Balta, D., M. Sellami, P. Kuhn, U. Schöpp, M. Buchinger, N. Baracaldo, A. Anwar, H. Ludwig, M. Sinn, M. Purcell, et al. (2021). "Accountable Federated Machine Learning in Government: Engineering and Management Insights". In: *International Conference on Electronic Participation*. Springer.

Bennett, C. J. (2016). "Voter databases, micro-targeting, and data protection law: can political parties campaign in Europe as they do in North America?" In: *International Data Privacy Law*.

Bonawitz, K., V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth (2017). "Practical secure aggregation for privacy-preserving machine learning". In: *ACM Conference on Computer & Communications Security*.

Chen, C. P. and C.-Y. Zhang (2014). "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data". In: *Information sciences*.

Chen, D. and H. Zhao (2012). "Data security and privacy protection issues in cloud computing". In: *2012 International Conference on Computer Science and Electronics Engineering*. IEEE.

Chen, Y., X. Qin, J. Wang, C. Yu, and W. Gao (2020). "Fedhealth: A federated transfer learning framework for wearable healthcare". In: *IEEE Intelligent Systems*.

Clarkson, G., T. E. Jacobsen, and A. L. Batcheller (2007). "Information asymmetry and information sharing". In: *Government Information Quarterly*.

Duan, Y., J. S. Edwards, and Y. K. Dwivedi (2019). "Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda". In: *International Journal of Information Management*.

General Data Protection Regulation (2018). European Commission. URL: https://gdpr.eu/.

Guberović, E., C. Alexopoulos, I. Bosnić, and I. Čavrak (2022). "Framework for Federated Learning Open Models in e-Government Applications". In: *Interdisciplinary Description of Complex Systems*.

Isaak, J. and M. J. Hanna (2018). "User data privacy: Facebook, Cambridge Analytica, and privacy protection". In: *Computer*.

Jiang, J. C., B. Kantarci, S. Oktug, and T. Soyata (2020). "Federated learning in smart city sensing: Challenges and opportunities". In: *Sensors*.

Johnson, P. A. (2016). "Reflecting on the success of open data: How municipal government evaluates their open data programs". In: *International Journal of E-Planning Research*.

Jordan, M. I. and T. M. Mitchell (2015). "Machine learning: Trends, perspectives, and prospects". In: *Science*.

Kang, J., Z. Xiong, D. Niyato, S. Xie, and J. Zhang (2019). "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory". In: *IEEE Internet of Things Journal*.

Lachana, Z., C. Alexopoulos, E. Loukis, and Y. Charalabidis (2018). "Identifying the different generations of Egovernment: an analysis framework". In: *The 12th Mediterranean Conference on Information Systems*.

Li, L., Y. Fan, M. Tse, and K.-Y. Lin (2020a). "A review of applications in federated learning". In: *Computers & Industrial Engineering*.

Li, Q., Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He (2021). "A survey on federated learning systems: vision, hype and reality for data privacy and protection". In: *IEEE Transactions on Knowledge and Data Engineering*.

Li, T., A. K. Sahu, A. Talwalkar, and V. Smith (2020b). "Federated learning: Challenges, methods, and future directions". In: *IEEE Signal Processing Magazine*.

Liu, X., S. Shao, Y. Yang, K. Wu, W. Yang, and H. Fang (2021). "Secure Federated Learning Model Verification: A Client-side Backdoor Triggered Watermarking Scheme". In: *IEEE International Conference on Systems, Man, and Cybernetics*. IEEE.

Liu, Y., J. James, J. Kang, D. Niyato, and S. Zhang (2020). "Privacy-preserving traffic flow prediction: A federated learning approach". In: *IEEE Internet of Things Journal*.

Manoj, T, K. Makkithaya, and V. Narendra (2022). "A Federated Learning-Based Crop Yield Prediction for Agricultural Production Risk Management". In: *2022 IEEE Delhi Section Conference*. IEEE.

McMahan, B., E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas (2017). "Communication-efficient learning of deep networks from decentralized data". In: *Artificial intelligence and statistics*. PMLR.

Mikhaylov, S. J., M. Esteve, and A. Campion (2018). "Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration". In: *Philosophical transactions of the royal society a: mathematical, physical and engineering sciences*.

Mitra, A., R. Jaafar, G. J. Pappas, and H. Hassani (2021). "Linear convergence in federated learning: Tackling client heterogeneity and sparse gradients". In: *Advances in Neural Information Processing Systems*.

Mothukuri, V., R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava (2021). "A survey on security and privacy of federated learning". In: *Future Generation Computer Systems*.

OECD (2019). *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies*.

Olson, M. (1965). "The logic of collective action Harvard University Press". In: *Cambridge, MA*.

Passerat-Palmbach, J., T. Farnan, M. McCoy, J. D. Harris, S. T. Manion, H. L. Flannery, and B. Gleim (2020). "Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data". In: IEEE.

Pingitore, G., V. Rao, K Dwivedi, and K Cavallaro (2017). "To share or not to share". In: URL: https://www2.deloitte.com/content/dam/insights/us/articles/4020_To-share-or-not-to-share/DUP_To-share-or-not-to-share.pdf.

Shae, Z. and J. Tsai (2018). "Transform blockchain into distributed parallel computing architecture for precision medicine". In: *International Conference on Distributed Computing Systems*. IEEE.

Truong, N., K. Sun, S. Wang, F. Guitton, and Y. Guo (2021). "Privacy preservation in federated learning: An insightful survey from the GDPR perspective". In: *Computers & Security*.

Tuor, T., S. Wang, B. J. Ko, C. Liu, and K. K. Leung (2021). "Overcoming noisy and irrelevant data in federated learning". In: *International Conference on Pattern Recognition*. IEEE.

Ward, B. T. and J. C. Sipior (2010). "The Internet jurisdiction risk of cloud computing". In: *Information systems management*.

Webster, J. and R. T. Watson (2002). “Analyzing the past to prepare for the future: Writing a literature review”. In: *MIS quarterly*.

Wheaton, S. and Martuscelli (2021). “WHO, Berlin float sanctions if countries suppress information on pandemics”. In: *Politico*. URL: https://www.politico.eu/article/who-berlin-float-sanctions-if-countries-suppress-information-on-pandemics/ (visited on 05/20/2022).

WHO (2021). *Global leaders unite in urgent call for international pandemic treaty*. URL: https://www.who.int/news/item/30-03-2021-global-leaders-unite-in-urgent-call-for-international-pandemic-treaty.

Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. (2016). “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific data*.

Wiseman, J. (2020). “Silo busting: The challenges and success factors for sharing intergovernmental data”. In: *IBM Center for The Business of Government. Accessed April* 6, p. 2021.

Xu, J., B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang (2021). “Federated learning for healthcare informatics”. In: *Journal of Healthcare Informatics Research*.

Yang, D., Z. Xu, W. Li, A. Myronenko, H. R. Roth, S. Harmon, S. Xu, B. Turkbey, E. Turkbey, X. Wang, et al. (2021). “Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan”. In: *Medical image analysis*.

Yang, Q., Y. Liu, T. Chen, and Y. Tong (2019). “Federated machine learning: Concept and applications”. In: *ACM Transactions on Intelligent Systems and Technology*.

Yu, H., Z. Liu, Y. Liu, T. Chen, M. Cong, X. Weng, D. Niyato, and Q. Yang (2020). “A fairness-aware incentive scheme for federated learning”. In: *Conference on AI, Ethics, and Society*.

Ziller, A., A. Trask, A. Lopardo, B. Szymkow, B. Wagner, E. Bluemke, J.-M. Nounahon, J. Passerat-Palmbach, K. Prakash, N. Rose, et al. (2021). “Pysyft: A library for easy federated learning”. In: *Federated Learning Systems*.

Research Paper 5 – *Federated Learning: Organizational Opportunities, Challenges, and Adoption Strategies*

Full Title:
Federated Learning: Organizational Opportunities, Challenges, and Adoption Strategies

Authors:
Joaquín Delgado Fernández, Martin Brennecke,Tom Josua Barbereau, Alexander Rieger, Gilbert Fridgen

Keywords:
Artificial intelligence, Deep learning, Federated learning, Organization theory

Under Review / Preprint: https://arxiv.org/pdf/2308.02219.pdf

Abstract

Restrictive rules for data sharing in many industries have led to the development of Federated Learning (FL). FL is a Machine Learning (ML) technique that allows distributed clients to train models collaboratively without the need to share their respective training data with others. In this article, we first explore the technical basics of FL and its potential applications. Second, we present a quadrant to map organizations along the lines of their artificial intelligence capabilities. We then discuss why different organizations in different industries, including industry consortia, established banks, public authorities, and data-intensive Small and midsize enterprises might consider different approaches to FL. To conclude, we argue that FL presents an institutional shift with ample research opportunities for the business and information systems engineering community.

1 Introduction

Artificial intelligence (AI) capabilities have become an important enabler in various industries (Berente et al., 2021). Open-source models and AI-as-a-service offerings have substantially reduced the costs of acquiring these capabilities, shifting the focus to calibrating these AI offerings and training the underlying models with the ’right’ data (Guntupalli and Rudramalla, 2023; Lins et al., 2021).

The challenge of data availability becomes apparent in consideration of the requirements under which Machine learning (ML) techniques can produce effective results: they are frequently attributed as “data-hungry” (Adadi, 2021). They are data-hungry because the underlying principle of these techniques is to automate the extraction of complex representations or abstractions manifested in data (Najafabadi et al., 2015). Data availability is here understood in terms of quantity and quality, both pivotal to improve performance (Fan, 2015). However, data of sufficient quantity and quality is not always available. This is because data is costly, requires advanced Information Technology (IT) capabilities (Aral and Weill, 2007), and is often restricted to organizational boundaries for competitive, regulatory, and ethical reasons (Adadi, 2021; Berente et al., 2021; Jordan and Mitchell, 2015) and especially when data needs to be shared across organizational boundaries (Ko et al., 2019).

Federated learning (FL) promises to mediate these data-sharing concerns. It allows organizations to cooperate in training a ML model without sharing data across organizational boundaries (Kalra et al., 2023). In particular, it allows organizations in data-driven environments to co-create value from data without compromising sensitive information and data privacy. Its areas of application range from financial services (Lee et al., 2023) to healthcare (Kaissis et al., 2021) and public administration (Kalra et al., 2023; Pati et al., 2022; Sprenkamp et al., 2023).

In this article, we explain how FL works on a conceptual level and how it differs from conventional/centralized approaches for the training of ML models. We then discuss the contexts in which FL can create organizational value before discussing the challenges that come with its implementation. Based on these challenges, we present open questions and opportunities for further technical, organizational, and legal research surrounding FL.

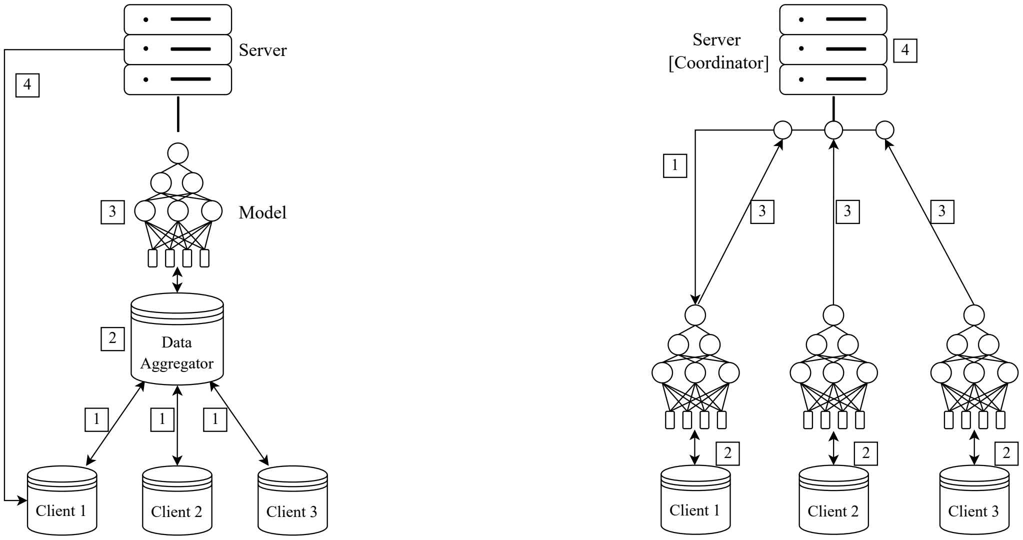
2 Technical Foundations

Federated Learning has its origins in projects at Google and OpenAI that sought to use data generated by mobile devices to train ML models that enhance user experiences. These projects quickly realized that much of this data was highly personal and sensitive, which complicated its upload to company servers (Abadi et al., 2016; McMahan et al., 2016). FL was introduced in 2016 to circumvent these complications and enable a "machine learning setting where the goal is to train a high-quality centralized model while training data remains distributed over a large number of clients, each with unreliable and relatively slow network connections" (Konečný et al., 2016). In the following years, the use of FL was extended to other areas, not least due to the ability of FL to support ML on data that is not independently and identically distributed (non-IID) (Zhao et al., 2018). More recent advances also eliminated the need for synchronous training and communication (Xu et al., 2021) and central servers for coordination of the decentralized learning process (Chang et al., 2018; Kalra et al., 2023; Shen et al., 2020, see).

The dynamic nature of FL makes it challenging to present a complete overview of all its approaches. In the following, we thus rather focus on the conceptual differences between FL and more conventional/centralized approaches to the training of ML models on cross-organizational data (Figure 1).

In conventional/centralized ML approaches, ML models are trained on (anonymized) data that is stored in a central repository. In more technical terms, the 'learning' process is the following:

1. Clients send their data (usually anonymized) to the central repository.



(a) Centralized.

(b) Decentralized/Federated.

Figure 1: Architectures for machine learning.

2. The central repository consolidates this data into a common format, homogenizing it across clients and datasets.

3. A pre-defined model is trained on the consolidated data.

4. The trained models are sent back to the clients that submitted their data.

In the case of FL, each client maintains control over its data and individual clients are responsible for preparing it for the training of shared ML models. Once the required data is prepared, the clients must establish a process to coordinate and streamline the training of the FL model. Once all clients have agreed to this process, they can start locally training a (partial) ML model while maintaining full control of their data. FL takes advantage of the distribution of clients by eliminating the need for data aggregation but rather aggregating (partial) ML models. Usually, this aggregation of (partial) ML models occurs on a central server. In more technical terms, the steps in the 'learning' process are the following:

1. A subset of clients downloads the most recent version of the 'global' ML model (typically from a central server).

2. Each of these clients trains an updated model using their local data.

3. Each client uploads their updated models (typically to the central server).
4. The ′local′ ML models are aggregated to build a better ′global′ model and the process resumes with the first step.

Multiple algorithms have been developed to select clients in each training round and aggregate their updated models to overcome problems with model heterogeneity, stability, and synchronization (see Zhang et al., 2021). The idea is always the same, however: ML models are trained locally, shared with the other clients or a central server, aggregated, and then re-routed to clients. These consecutive steps can be repeated until a certain performance benchmark or a set number of training rounds have been reached.

3 Organizational Considerations for the Use of Federated Learning

Attractive FL applications will typically be situated in a triangle between data regulation, data competition, and AI capabilities (see Figure2). While data regulation and competition place limits on data sharing, AI capabilities define the degree to which organizations can successfully train, deploy and use their own ML models.

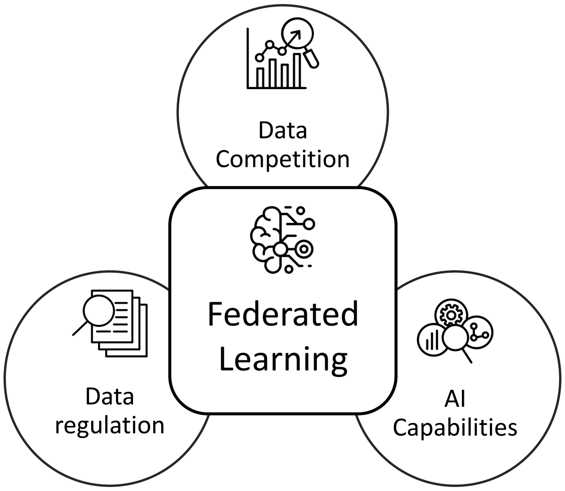


Figure 2: Conceptualization of the interplay between data regulation, data competition, and AI capabilities.

3.1 Data Regulation

In many industries, regulation places strict limits on cross- and intra-organizational data sharing and - by extension - on the collaborative training of ML models (Berente et al., 2021; Buxmann et al., 2021). In Europe, for example, the General Data Protection Regulation (GDPR), the Data Act, and the AI Act restrict how organizations are able to share personal and non-personal data. The result of these regulatory restrictions are often small(er) datasets for training and ML models with low(er) performance (Brauneck et al., 2023).

For organizations required to comply with these and similar regulations, FL can be highly valuable. As FL enables the decentralized training and subsequent sharing of models rather than their underlying training data, it reduces regulatory risks related to data sharing. FL can further provide substantial benefits to public authorities, which are often required to maintain decentralized databases and registries and not to share data unless there is a legal basis (Sprenkamp et al., 2023).

3.2 Data Competition

A second dimension for the use of FL is the prevalence of data-based competition in the adopting organization's industry. Organizations in such environments are often reluctant to share data, especially when data control comes with control over the appropriation of data network effects (Abbas et al., 2021; Gregory et al., 2021; Leiponen, 2002).

FL may enable these organizations to cooperate without conflicts about data control. It may also enable them to define fair schemes for the appropriation of data network effects based on the 'value' each training dataset contributes to the joint ML model. FL may be especially attractive to small and medium-sized organizations that can use FL to compete with larger organizations (Mazzocca et al., 2023; Zhang et al., 2023). But it may also be attractive to large organizations as it allows them to promote de-facto standards for the use of ML models in their industry and how value can be appropriated from their use providing abstractions of complex systems at multiple levels or from different viewpoints (Mohagheghi et al., 2013).

3.3 AI capabilities

The third dimension that organizations interested in FL should consider relates to their abilities to design, train, and run ML models. Despite the increasing prevalence of open-source ML models and AI-as-a-service offerings (Lins et al., 2021), many organizations struggle with the capabilities required to make use of these models and offerings (Enholm et al., 2021; Lins et al., 2021). The challenges reach from limited technical capabilities to a lack of personnel that can translate between business departments and AI engineers.

FL can alleviate some of these challenges by not only pooling data but also the capabilities required to design, train, and run ML models. Especially for organizations with weaker technological capabilities, it can make sense to partner with and learn from stronger organizations. Yet also those with stronger technological capabilities stand to benefit when partner organizations can add superior business capabilities.

4 Adoption challenges

Building on the considerations in Section 3, we now turn to the adoption challenges associated with FL. To structure our discussion of these challenges, we employ a simple framework to distinguish organizations interested in FL according to the degree of competition in their industry and the level of their AI capabilities (see Figure 3).

4.1 Type 1 - Strong AI Capabilities & Low Competition

Type 1 organizations have notable AI capabilities and operate in environments with low competition. They can profit from FL when there are substantial benefits from highly accurate models but regulation complicates data sharing. A good example is medical R&D consortia (Bi et al., 2023; Malin et al., 2013).

The central concern for Type 1 organizations will typically be security and data protection. Although FL promises high security and data protection abilities, it cannot guarantee them (Benmalek et al., 2022). More specifically, the security of FL can be compromised by attacks on the communication channels between the clients (Chatterjee et al., 2020; Wang et al., 2020). This attack vector is difficult to eliminate but it can

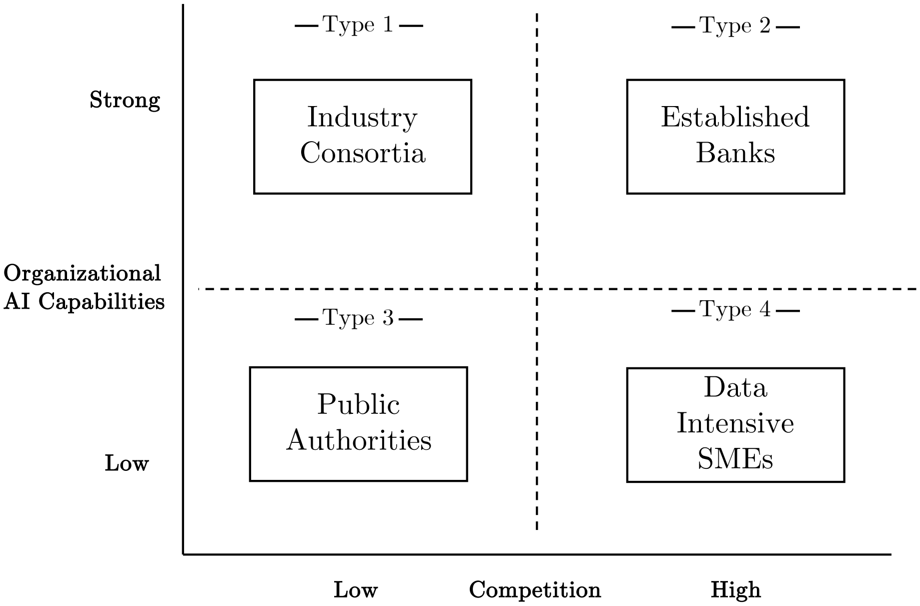


Figure 3: A conceptual framework for the adoption of FL

be mediated by techniques such as secure multiparty computation (SMPC) and secure aggregation (SecAgg) (Bonawitz et al., 2021; Kaissis et al., 2021). Data protection, in turn, can be compromised when training data can be reverse-engineered from model updates (see Bagdasaryan et al., 2020; Shejwalkar and Houmansadr, 2021). Solutions to this challenge include Differential Privacy (DP), which adds noise to either the training data or the model updates before they are shared with other clients (Dwork, 2006). While the use of SecAgg does not appear to significantly reduce the overall performance of a model compared to the use of training a regular FL model, the use of DP has been shown to have a stronger impact on model performance (Fernández et al., 2022; Kaissis et al., 2021) due to the addition of noise to the model updates (McMahan et al., 2018).

4.2 Type 2 - Strong AI Capabilities & High Competition

Type 2 organizations are characterized by comparatively strong AI capabilities and a highly competitive environment. They stand to benefit from FL by monetizing their own capabilities and know-how while gaining insights from more agile competitors or when regulation complicates the sharing of data between sister companies. Examples include established financial institutions or multi-national data brokers.

Despite their strong AI capabilities, implementing FL can be challenging for Type 2 organizations. The first challenge relates to establishing set-ups that do not violate antitrust laws. This challenge is especially acute when several large organizations collaborate or have an outsized influence in joint FL projects. One way of addressing this challenge could be to open-source the final model or create an auditable data trust, thereby reducing the risk of market power abuse (Mahari et al., 2021).

Second, Type 2 organizations might consider working with sister companies. In intra-organizational data sharing and cooperation, data-sharing restrictions might also apply, for example in cases where one unit of an organization must have a clear separation from another part of an organization. While other regulatory requirements might still apply, FL could allow these organizations, to share knowledge without sharing their respective underlying data with each other. Despite their high AI capabilities, Type 2 organizations might nevertheless struggle to address challenges such as the heterogeneity of data, as different data structures, formats, and distributions may exist. Combined with a wide variety of infrastructure, the use of FL poses significant technical challenges. As a result, maintaining a well-performing and useful model across all affiliates can be difficult to coordinate and manage.

4.3 Type 3 - Low AI Capabilities & Low Competition

Type 3 organizations are characterized by comparatively weak AI capabilities and operate in environments with low competition. They stand to benefit from using FL by ′pooling′ insights from data as well as the capabilities required to successfully design, use, and train ML models. Public authorities are one example of Type 3 organizations and often lack the capabilities to use their limited data effectively (Sprenkamp et al., 2023). These organizations also face challenges in internalizing external capabilities that would enable them to adopt new technologies. Generally, their challenges are driven by limitations in AI know-how acquisition, internal technical expertise, digital debt (Rolland et al., 2018),and personal data sharing (Isaak and Hanna, 2018).

FL-adoption in Type 3 organizations is often complicated by legacy information management systems and inadequate data preparation, the repeated delays that occur in the development of internal capabilities, as well as the high costs associated with acquiring external capabilities (Kuziemski and Misuraca, 2020). Type 3 organizations may also lack the expertise and resources required to fully leverage FL and its associated benefits, which impedes widespread adoption. To overcome these limitations, they might rely on IT service providers, which create external dependencies.

As Type 3 organizations regularly lack the financial and technical resources to rapidly build up internal AI capabilities or hire external IT service providers on a long-term basis, they might be less inclined to fully embrace new technologies, including FL. Additionally, and especially in the case of public authorities, Type 3 organizations may also face legal restrictions on the use and processing of data (Yang and Wu, 2020).

4.4 Type 4 - Low AI Capabilities & High Competition

Type 4 organizations are characterized by comparatively weak AI capabilities and a highly competitive environment. They stand to benefit from FL as it allows them to pool their limited AI capabilities with other organizations facing similar challenges (Lee et al., 2023) but also through learning effects created in collaboration with more advanced organizations. Examples of Type 4 organizations include data-intensive small and medium-sized enterprise (SME), such as new banks and fintechs startups.

Type 4 organizations have the potential to benefit from participation in FL implementations. However, the long-term success of such collaborations in highly competitive environments hinges on appropriate governance structures. Setting up these structures, in turn, involves organizational, economic, and legal considerations that must enable the effective management of the involved parties. While organizational trade-offs often revolve around losing agency and control to gain access to FL, economic incentives and compensations must be aligned to encourage organizational participation. Regarding the legal dimensions, clear agreements have to be made that adequately address intellectual property concerns and safeguard the interests of all parties involved. Overcoming these challenges is essential to the adoption and implementation of FL in Type 4 organizations.

5 Research Opportunities

Cross-organizational machine learning projects often face significant challenges due to regulatory or competitive limits to data sharing. FL provides a promising solution for these projects because it enables organizations to collaborative train FL models without sharing training data directly, fostering a more privacy-preserving environment and unlocking new opportunities for innovation and problem-solving in cross-organizational ML initiatives. As organizations seek to explore these new opportunities, some commonly held beliefs may be questioned and require reconsideration. We present technical, organizational, and legal research opportunities to inspire future scientific inquiries into the opportunities and challenges surrounding FL.

The first set of research opportunities and questions pertains to the *technical characteristics* and applications of FL (see Lins et al., 2021). Despite continuous improvements in the performance of specific FL implementations, organizations still hold reservations against the use of FL, due to a perceived technical immaturity. While solutions concerning the performance and security of selected FL models are to be developed within more technical disciplines, and how beyond the technical realm, there are barriers to adoption. Related questions for information systems researchers at the intersection of business and technology could be:

- What are the best practices to ensure privacy in a FL setting?
- How can FL be leveraged in real-time decision-making?
- How can FL models be protected from the risks posed by malevolent clients?

The second set of research opportunities and questions pertains to the *organizational contexts* in which FL is used and how AI is managed therein (see Berente et al., 2021). As FL enables the inclusion of numerous decentralized parties in a collaborative setting, its governance requirements differ considerably from those of conventional/centralized ML development and usage. While we expect that many insights into the organizational governance of conventional/centralized ML also hold for FL, cross-organizational governance frameworks might have to be adapted to FL operations. Along these lines, possible research questions for information systems researchers investigating the intersection of organizations and technology are:

- What server-level standards need to be set to reach the level of trust required for entities to partake in FL?

- What governance structures are needed for server-centric and server-less FL?
- What is the value of FL within different industries?
- How can FL be used to improve business processes and public services?
- What business models enable the commercialization of FL?
- How can FL applications be standardized for collaboration and commercialization?

The third set of research opportunities pertains to the *regulatory compliance* of specific FL implementations in various jurisdictions (see Truong et al., 2021). Although recent developments in FL enable organizations to resolve tensions between regulatory compliance and the benefits of ML, these developments also create new regulatory uncertainties, and it remains to be seen how regulators and organizations will position themselves in terms of FL-regulation. Along these lines, possible research questions for information systems researchers investigating the intersection of law and technology are:

- How can organizations ensure that their FL implementation is privacy-preserving?
- What are the legal and regulatory implications of FL?
- Under which regulatory conditions could successful and unbiased FL models flourish?

How a given FL system, particularly in cross-organizational settings, is governed directly impacts its adoption. Information systems researchers may draw on past experiences based on the deployment and adoption of distributed systems (Liang et al., 2021) and business process modeling (Recker et al., 2009) to contribute to a comprehensive understanding of this relationship. Regarding technical characteristics in specific, design science research (Hevner et al., 2004) is also well-positioned to develop and evaluate specific FL implementations and their sustained value creation over time.

6 Conclusion

In this article on FL, we describe how restrictive rules on data sharing and privacy resulted in the development of FL, an ML technique that enables organizations to collaboratively train models in a decentralized manner. As a basis for subsequent analysis, we summarize the technical foundations of FL and how it differs from conventional/centralized ML approaches. We then discuss the challenges of adopting FL for four different types of organizations, taking into account organizational requirements related to data-sharing regulations, market conditions, and technical developments. These four types of organizations operate in environments with either high or low levels of competition and have either high or low AI capabilities. We find that organizations consider using FL for different reasons and face varying challenges in its adoption, including the use of the technology itself, its regulatory environment, and its governance. FL has the potential to revolutionize the adoption and use of AI as it simultaneously addresses data privacy and capability acquisition concerns. It also has the potential to expand the AI capabilities of an organization beyond its boundaries.

Credit authorship contribution statement

Conceptualization, J.D.F, T.B; Methodology, J.D.F, T.B, A.R; Writing - Original Draft, J.D.F, T.B, M.B.; Supervision A.R, G.F.; Writing - Review & Editing, A.R, G.F.; Visualization, J.D.F; Funding acquisition, G.F. All authors have read and agreed to the published version of the manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

We would like to acknowledge Chul Min Lee for his valuable feedback on the first draft of this paper.

Moreover, we would like to acknowledge the support of the European Union (EU) within its Horizon 2020 programme, project MDOT (Medical Device Obligations Taskforce), Grant agreement 814654. Additionally, supported by Banque et caisse d'épargne de l'État, Luxembourg (Spuerkeess) and by PayPal together with the Luxembourg National Research Fund FNR (P17/IS/13342933/PayPal-FNR/Chair in DFS/Gilbert Fridgen).

References

Abadi, M., A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang (2016). "Deep learning with differential privacy". In: *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318.

Abbas, A. E., W. Agahari, M. Van de Ven, A. Zuiderwijk, and M. De Reuver (2021). "Business data sharing through data marketplaces: A systematic literature review". In: *Journal of Theoretical and Applied Electronic Commerce Research* 16.7, pp. 3321–3339.

Adadi, A. (2021). "A survey on data-efficient algorithms in big data era". In: *Journal of Big Data* 8. DOI: 10.1186/s40537-021-00419-9.

Aral, S. and P. Weill (2007). "IT assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation". In: *Organization science* 18.5, pp. 763–780.

Bagdasaryan, E., A. Veit, Y. Hua, D. Estrin, and V. Shmatikov (2020). "How to backdoor federated learning". In: *International conference on artificial intelligence and statistics*. PMLR, pp. 2938–2948.

Benmalek, M., M. A. Benrekia, and Y. Challal (2022). "Security of federated learning: Attacks, defensive mechanisms, and challenges". In: *Revue des Sciences et Technologies de l'Information-Série RIA: Revue d'Intelligence Artificielle* 36.1, pp. 49–59.

Berente, N., B. Gu, J. Recker, and R. Santhanam (2021). "Managing Artifical Intelligence". In: *MIS Quarterly* 45.3.

Bi, X., A. Gupta, and M. Yang (2023). *Understanding Partnership Formation and Repeated Contributions in Federated Learning: An Analytical Investigation*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.3986446. URL: https://papers.ssrn.com/abstract=3986446 (visited on 07/26/2023).

Bonawitz, K., P. Kairouz, B. McMahan, and D. Ramage (2021). "Federated Learning and Privacy: Building Privacy-Preserving Systems for Machine Learning and Data Science on Decentralized Data". In: *Queue* 19.5, 87–114. URL: https://doi.org/10.1145/3494834.3500240.

Brauneck, A., L. Schmalhorst, M. M. Kazemi Majdabadi, M. Bakhtiari, U. Völker, C. C. Saak, J. Baumbach, L. Baumbach, and G. Buchholtz (2023). "Federated machine learning in data-protection-compliant research". In: *Nature Machine Intelligence* 5, pp. 2–4. ISSN: 2522-5839. DOI: 10.1038/s42256-022-00601-5.

Buxmann, P., T. Hess, and J. B. Thatcher (2021). "AI-based information systems". In: *Business & Information Systems Engineering* 63.1, pp. 1–4.

Chang, K., N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D. Rubin, and J. Kalpathy-Cramer (2018). "Distributed deep learning networks among institutions for medical imaging". In: *Journal of the American Medical Informatics Association : JAMIA* 25. DOI: 10.1093/jamia/ocy017.

Chatterjee, P., E. Benoist, and A. Nath (2020). *Applied approach to privacy and security for the Internet of things*. IGI Global.

Dwork, C. (2006). "Differential Privacy". In: *Automata, Languages and Programming*. Ed. by M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–12.

Enholm, I. M., E. Papagiannidis, P. Mikalef, and J. Krogstie (2021). "Artificial Intelligence and Business Value: a Literature Review". In: *Information Systems Frontiers* 24.5, pp. 1709–1734. DOI: 10.1007/s10796-021-10186-w. URL: https://doi.org/10.1007/s10796-021-10186-w.

Fan, W. (2015). "Data quality: From theory to practice". In: *Acm Sigmod Record* 44.3, pp. 7–18.

Fernández, J. D., S. P. Menci, C. M. Lee, A. Rieger, and G. Fridgen (2022). "Privacy-preserving federated learning for residential short-term load forecasting". In: *Applied Energy* 326, p. 119915. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2022.119915. URL: https://www.sciencedirect.com/science/article/pii/S0306261922011722.

Gregory, R. W., O. Henfridsson, E. Kaganer, and H. Kyriakou (2021). "The Role of Artificial Intelligence and Data Network Effects for Creating User Value". In: *Academy of Management Review* 46.3, pp. 534–551. DOI: 10.5465/amr.2019.0178. URL: https://doi.org/10.5465/amr.2019.0178.

Guntupalli, N. and V. Rudramalla (2023). "Artificial Intelligence as a Service: Providing Integrity and Confidentiality". In: *Multi-disciplinary Trends in Artificial Intelligence*. Ed. by R. Morusupalli, T. S. Dandibhotla, V. V. Atluri, D. Windridge, P. Lingras, and V. R. Komati. Cham: Springer Nature Switzerland, pp. 309–315.

Hevner, A. R., S. T. March, J. Park, and S. Ram (2004). "Design science in information systems research". In: *MIS Quarterly*, pp. 75–105.

Isaak, J. and M. J. Hanna (2018). "User data privacy: Facebook, Cambridge Analytica, and privacy protection". In: *Computer*.

Jordan, M. I. and T. M. Mitchell (2015). "Machine learning: Trends, perspectives, and prospects". In: *Science*.

Kaissis, G., A. Ziller, J. Passerat-Palmbach, T. Ryffel, D. Usynin, A. Trask, I. Lima, J. Mancuso, F. Jungmann, M.-M. Steinborn, et al. (2021). "End-to-end privacy preserving deep learning on multi-institutional medical imaging". In: *Nature Machine Intelligence* 3.6, pp. 473–484.

Kalra, S., J. Wen, J. C. Cresswell, M. Volkovs, and H. R. Tizhoosh (2023). "Decentralized federated learning through proxy model sharing". en. In: *Nature Communications* 14.1.

Ko, B., S. Wang, T. He, and D. Conway-Jones (2019). "On Data Summarization for Machine Learning in Multi-organization Federations". In: *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 63–68. URL: https://doi.org/10.1109/SMARTCOMP.2019.00030.

Konečný, J., B. McMahan, D. Ramage, and P. Richtárik (2016). *Federated Optimization: Distributed Machine Learning for On-Device Intelligence*. DOI: 10.48550/ARXIV.1610.05492. arXiv: 1610.02527.

Kuziemski, M. and G. Misuraca (2020). "AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings". In: *Telecommunications policy* 44.6, p. 101976.

Lee, C. M., J. Delgado Fernandez, S. Potenciano Menci, A. Rieger, and G. Fridgen (2023). "Federated Learning for Credit Risk Assessment". In: *Proceedings of the 56th Hawaii International Conference on System Sciences*.

Leiponen, A. E. (2002). "Why Do Firms Not Collaborate? The Role of Competencies and Technological Regimes". In: *Innovation and Firm Performance*.

Liang, T.-P., R. Kohli, H.-C. Huang, and Z.-L. Li (2021). “What drives the adoption of the blockchain technology? A fit-viability perspective”. In: *Journal of Management Information Systems* 38.2, pp. 314–337.

Lins, S., K. D. Pandl, H. Teigeler, S. Thiebes, C. Bayer, and A. Sunyaev (2021). “Artificial Intelligence as a Service”. In: *Business & Information Systems Engineering* 63.4, pp. 441–456.

Mahari, R., S. C. Lera, and A. Pentland (2021). *Time for a new antitrust era: refocusing antitrust law to invigorate competition in the 21st century*.

Malin, B. A., K. E. Emam, and C. M. O’Keefe (2013). “Biomedical data privacy: problems, perspectives, and recent advances”. In: *Journal of the American Medical Informatics Association : JAMIA* 20 1, pp. 2–6. URL: https://doi.org/10.1136/amiajnl-2012-001509.

Mazzocca, C., N. Romandini, M. Mendula, R. Montanari, and P. Bellavista (2023). “TruFLaaS: Trustworthy Federated Learning as a Service”. In: *IEEE Internet of Things Journal*, pp. 1–1. DOI: 10.1109/JIOT.2023.3282899.

McMahan, B., E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas (2016). *Communication-Efficient Learning of Deep Networks from Decentralized Data*. DOI: 10.48550/ARXIV.1602.05629.

McMahan, B., D. Ramage, K. Talwar, and L. Zhang (2018). *Learning Differentially Private Recurrent Language Models*. arXiv: 1710.06963 [cs.LG].

Mohagheghi, P., W. Gilani, A. Stefanescu, M. A. Fernandez, B. Nordmoen, and M. Fritzsche (2013). “Where does model-driven engineering help? Experiences from three industrial cases”. In: *Software & Systems Modeling* 12, pp. 619–639.

Najafabadi, M., F. Villanustre, T. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic (2015). “Deep learning applications and challenges in big data analytics”. In: *Journal of Big Data* 2. DOI: 10.1186/s40537-014-0007-7.

Pati, S., U. Baid, B. Edwards, M. Sheller, S.-H. Wang, G. A. Reina, P. Foley, A. Gruzdev, D. Karkada, C. Davatzikos, et al. (2022). “Federated Learning Enables Big Data for Rare Cancer Boundary Detection”. In: *arXiv preprint arXiv:2204.10836*.

Recker, J., M. Rosemann, M. Indulska, and P. Green (2009). “Business process modeling-a comparative analysis”. In: *Journal of the Association for Information Systems* 10.4, p. 1.

Rolland, K. H., L. Mathiassen, and A. Rai (2018). "Managing digital platforms in user organizations: The interactions between digital options and digital debt". In: *Information Systems Research* 29.2, pp. 419–443.

Shejwalkar, V. and A. Houmansadr (2021). "Manipulating the byzantine: Optimizing model poisoning attacks and defenses for federated learning". In: *NDSS*.

Shen, T., J. Zhang, X. Jia, F. Zhang, G. Huang, P. Zhou, K. Kuang, F. Wu, and C. Wu (2020). *Federated Mutual Learning*. arXiv: 2006.16765 [cs.LG].

Sprenkamp, K., J. Delgado Fernandez, S. Eckhardt, and L. Zavolokina (2023). "Federated Learning as a Solution for Problems Related to Intergovernmental Data Sharing". In: *Proceedings of the 56th Hawaii International Conference on System Sciences*, p. 10. DOI: https://hdl.handle.net/10125/102838.

Truong, N., K. Sun, S. Wang, F. Guitton, and Y. Guo (2021). "Privacy preservation in federated learning: An insightful survey from the GDPR perspective". In: *Computers& Security* 110, p. 102402. ISSN: 0167-4048.

Wang, D., C. Li, S. Wen, S. Nepal, and Y. Xiang (2020). "Man-in-the-middle attacks against machine learning classifiers via malicious generative models". In: *IEEE Transactions on Dependable and Secure Computing* 18.5, pp. 2074–2087.

Xu, C., Y. Qu, Y. Xiang, and L. Gao (2021). "Asynchronous federated learning on heterogeneous devices: A survey". In: *arXiv preprint arXiv:2109.04269*.

Yang, T.-M. and M.-C. Wu (2020). "An Exploration of Factors Influencing Taiwan Government Agencies' Open Data Participation: A Multi-Group Analysis Perspective". In: *The 21st Annual International Conference on Digital Government Research*, pp. 356–358.

Zhang, C., Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao (2021). "A survey on federated learning". In: *Knowledge-Based Systems* 216, p. 106775.

Zhang, J., C. Cooper, and R. X. Gao (2023). "Federated Learning for Privacy-Preserving Collaboration in Smart Manufacturing". In: *Manufacturing Driving Circular Economy*. Ed. by H. Kohl, G. Seliger, and F. Dietrich. Cham: Springer International Publishing, pp. 845–853. ISBN: 978-3-031-28839-5.

Zhao, Y., M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra (2018). "Federated learning with non-iid data". In: *arXiv preprint arXiv:1806.00582*.