DOI: 10.15439/2023F5947
ISSN 2300-5963 ACSIS, Vol. 36

Federated Learning for Data Trust in Logistics

Michael Koch
Institute for applied Computer Science (InfAI) e.V.
Goerdelerring 9, 04109 Leipzig, Germany
Email: koch@infai.org

Sascha Kober, Stanislaw Straburzynski, Benjamin Gaunitz, Bogdan Franczyk
Leipzig University
Faculty of Economics
Grimmaische Straße 12
04109 Leipzig, Germany
Email: {kober, straburzynski, gaunitz, franczyk}@wifa.uni-leipzig.de

***Abstract*—In the field of logistics, there is a significant shortage of qualified employees. Artificial Intelligence (AI) can help solve that problem supporting existing employees and reducing their workload. However, large amounts of data to train AI models are required and, in most cases, due to lack of trust between companies, model training is based solely on locally stored data from logistics providers and some publicly available datasets. To address this data scarcity issue, a proposed solution is to employ federated learning (FL), in the context of data trust (DT) by training AI models across multiple companies, based on both centralized data, within the DT platform and decentralized data from logistics providers data silos, while ensuring data sharing access at the attribute level. This paper proposes this approach and points out the importance of data sharing for effective model training for solving workforce challenges in logistics.**

***Index Terms*— data trust, federated learning, logistics, machine learning, artificial intelligence**

I. Introduction

According to Germany's domestic freight transport statistics in 2021, 37.6% of transport vehicles were empty at every driven kilometer, a recurring trend as in previous years [1]. Additionally, the degree of utilization of loading capacity and transport performance has decreased from 40% in 2002 to 34.7% in 2021 and 45.9% to 41.4% respectively [1]. The degree of utilization of loading capacity is defined as how full a vehicle is in relation to its total loading capacity. The transport performance is a statistic that summarizes various key figures for freight transport. At the European Union level, one in every five kilometers is travelled by an empty vehicle in 2020 and an average of 24% of national transport is empty [2], [3]. Another problem is the shortage of drivers and logistics workers with high percentages of unfilled positions in various roles [4]. The logistics industry also faces issues related to communication, collaboration, flexibility in capacity planning [5] and the lack of reliance in existing online platforms [6].

To address these challenges, the use of data trust (DT) platforms as secure and transparent platforms for data exchange and storage [6] has been proposed. This DT is a neutral entity that serves a DT ecosystem for data exchange and is managed by a transparent and non-profit organization. This DT should implement state-of-the-art data security techniques that meet the requirements of logistics companies, including efficient access and usage control concepts at the attribute level, based on the logistics business data. This paper builds upon the ongoing TRANSIT project at the University of Leipzig in exchange with the participating logistics providers [7], as well as on research in current freight exchange platforms [8].

The idea is to support logistics service providers with artificial intelligence (AI) in the context of a DT platform. This can be done by providing incentives or creating a benefit for platform usage and thus addressing the shortage of qualified workers [9]. Examples of AI support include assisting workers in calculating transport prices, which is currently a complex process or helping them to find potential cooperation partners, because many of these processes are manual and based on the knowledge of just a few employees within a company. To achieve accurate results, a substantial amount of business data from each logistics provider is necessary [10]. However, this data are primarily stored in the local data silos of the logistics company rather than the DT platform. However, many companies prefer not to upload data to DT platform. Data minimization in

The position paper is funded by Federal Ministry of Education and Research (Project ID: 16DTM109B) and by the European Union - NextGenerationEU.

the General Data Protection Regulation (GPDR) [11] and potentially much larger historical data can be among the reasons.

Leveraging this vast source of data, without transporting and transforming it within the platform, is a challenge, but it increases the trust of logistics companies. The utilization of federated learning (FL) and efficient usage control for the external data stores provides a solution within a DT [12]. This solution trains the models in each of the data silos. Therefore, a comparable input format is required. Then the data trustee will act as a neutral authority at this point, advising the logistician on how to transform the data locally to obtain high-quality training data for any training purpose. The combination of DT and FL has the following advantages:

- Reducing the effort to merging, combining, and normalizing the local heterogeneous data
- High scalability on distributed and heterogeneous hardware
- Enhanced security, as there is no direct access to the locally stored data
- Enrichment of the data with publicly available data

To leverage this large amount of data and achieve a critical mass of users, it is crucial to train AI models securely and fairly [13]. A significant research topic in FL is ensuring fairness in the returned results obtained by users based on the quantity and quality of data used for training the model, as well as its impact on global model parameters and gradients. This idea is particularly relevant to the logistics sector, where established companies do not want to share the information which they have collected over the years and which shows some internals, like the price calculation.

Therefore, an effective mechanism for training and maintaining the model is important when considering scenarios where start-ups want to participate with their limited data or attackers want to obtain a data leak that reveals business secrets. For this proposal, the central global model should not predict exact values. Also, the parameters of the shared model should be noisy so that it is impossible to obtain data used during model training. Furthermore, the predictions and the downloaded model should be adjusted in accuracy based on the amount of data that a company has provided to the model training.

By adopting this approach, the platform can increase its usage, efficiently and effectively address the shortage of skilled workers, and facilitate better management and collaboration. The utilization of federated AI to consider various aspects of the market through a combined overall model, such as preventing price dumping by calculating fair market prices, provides benefits to all companies and will increase the usage of the platform.

II. Related Work and Background

This section provides an overview of related work and research that impacts on successful realization of the proposed solution. Topics which will be related to problem-solving are DT platforms, access, and usage control and at least FL.

A. *Data trust*

The concept of Data Trust (DT) has been introduced as a means of facilitating the exchange of data between different entities [6]. This DT can be imagined as a neutral entity, such as a not-for-profit organization or a transparent company, that establishes a data sharing ecosystem. This instance provides the infrastructure for data sharing, which can be designed in different ways, such as DT as a service [13].

The actors defined in the context of DT are the data trustee, data provider and data user. The data trustee is responsible for providing the above-mentioned infrastructure. The data provider, which can be a company or person, contributes the data for sharing. This can involve transmitting the data to the platform and storing it there or transmitting it directly to the data user while the metadata from the data source are stored securely on the DT.

This data user, represents persons and companies that utilize the provided data to develop innovative products for the data provider or other clients. However, they are not authorized to sell the data without the permission of the data provider, and they can only process the data on the terms and conditions agreed with the data provider.

Managing a DT involves considering a number of aspects, including internal governance, user interaction and market structure, which are introduced in [14]. Legal regulations, such as GDPR and Data Governance Act (DGA), also are related to this issue when managing data within a platform or as part of a DT [15], [11]. One approach, introduced by Lomotey et al. [13] is DT as a Service, wherein the data trustee deploys a DT platform for a specific application domain and integrates services into the platform on demand.

It is important to note that the data trustee does not generate revenue by handling the data. Instead, the DT finances itself through user license fees or providing additional services in consultation with the data providers.

B. *Access and Usage Control*

Access and usage control are essential for ensuring data security within the platform. This fine-grained access control ensures that data providers can define access permissions at the attribute level of entities, such as the street of an address. It should then support centralized and decentralized scenarios, to enable secure data sharing between logistic service providers and researchers. To fulfill these requirements, access and usage control concepts need to be

TABLE I.
ACCESS CONTROL

concept	short description	Source
IBAC	identity-based	[16]
RBAC	role-based	[17]
ABAC	attribute-based	[16], [18]
ReBAC	relation-based	[19]
UCON	usage control	[20]

analyzed, evaluated, and continuously developed at the conceptual and implementation levels. One recommended access control architecture for the underlying DT platform is the zero-trust model [16] where each data access request is thoroughly evaluated and can be combined with different access control mechanisms.

Table 1 provides an overview of existing access control concepts. Through preliminary research, literature review and analysis of existing software, it became apparent that a specific problem could not be solved directly and required workarounds. A major logistics requirement, which relates to setting permissions at the attribute level, is often addressed by mapping at the entity level, where a unique identifier exists. In addition, certain parts of the data may need to be shared with all platform participants, such as when seeking collaboration partners, while ensuring that certain government agencies have full access to a company's data. It is additionally important for the data provider to be able to distinguish and configure which individuals, or companies can access their data.

Furthermore, in the context of implementation of artificial intelligence in the DT platform, there is a need for use control to regulate how data can be processed. In this way, it is possible for all parties involved in logistics to determine which models can use their data for training purposes. The preferred access model is Attribute-Based Access Control (ABAC) because it can also incorporate Identity-Based Access Control (IBAC) and Role-Based Access Control (RBAC) as attributes of a sign-on user, providing a higher level of granularity. Relation-based access control (ReBAC) and usage control (UCON) are suitable for more complex use cases and are not currently required for this DT implementation. The advantage of using ReBAC is to define access on an flexible data model [17] and allow clean and fast retrieval of access rights, but in actual implementation it needs a unique identifier for all entities, which is not always present in logistics data exchange, as they also share data at the attribute level.

C. Federated Learning

FL is an approach to AI in which the data are kept in local data silos, and the algorithm is applied to those silos for training purposes. However, the data must be transformed locally into a comparable input format before the training. The trained models can then be merged either centrally or in a decentralized manner. This concept was first introduced by Mahan et al. [18] and involves multiple iterative steps in the training process including client selection, client computation, model aggregation, model update and convergence checking.

In the client selection step, the coordinator or an algorithm in the P2P network chose clients based on various criteria. These criteria include historical activity, such as previous involvement in the training process or computation time, as well as factors such as model quality, influence on the global model, quality of training data, and technical characteristics such as network bandwidth and memory.

The calculation step is then executed on the selected clients. They receive the global model trained in previous training rounds or initialized with gradients and execute training with their local data and the specified computational parameters.

The next step involves aggregating of the locally trained models. Multiple approaches for model aggregation exist, which may include verification of the model on cybersecurity, etc. [23]. To address privacy concerns, concepts such as Differential Privacy (DP) or Homomorphic Encryption (HE) can be applied before transmitting the local model [24]. On the one hand, this approach provides the advantage of securing locally sensitive data. On the other hand, there are challenges such as longer convergence times in certain cases of high privacy with DP and increased computational overhead using HE. In addition, there is a commonly used approach where not all selected clients need to send their updated model; only the majority of clients participate in the aggregation.

In the fourth step, the coordinator or selected clients in a decentralized environment updates the model. A variety of algorithms can be used for this update. These include gradient averaging and optimization techniques such as the Adam optimizer [24]. The aim is to speed up the convergence process and to improve the accuracy of the resulting model. It is also beneficial to adopt a secure aggregation protocol to enhance cyber security [24], but at the cost of increased communication overhead.

The last step is to verify that the convergence criteria have been met, typically by evaluating whether the convergence error is below a predefined threshold. The purpose of this step is to ensure that the training process has achieved a satisfactory level of accuracy of the resulting model. If the algorithm does not converge, it restarts from the selection step.

In the field of FL, recent studies have identified four main research directions [10], [25], [26]: cybersecurity, fairness, optimization of aggregation and computation of heterogeneous FL [10] and the analysis of these points in the context of blockchain technology [27].

Attempts have been made to combine these research directions and design a comprehensive framework or FL algorithm that addresses openness, security, fairness, and decentralization. Such a successful attempt can be found in

[28], which entails further testing and analysis in logistics assuming a DT environment.

Recent directions in FL explore centralized pre-training of models to improve convergence and accuracy [29]. In this setting, proxy data do not present the same privacy concerns as in a DT scenario, so that privacy mechanisms such as DP must also be applied at this point. Adaptive central training methods in FL for faster convergence have also been proposed [30]. Furthermore, there is ongoing research on the combination of FL and central learning, which is called mixed FL (MFL) [31], [32]. Until now, MFL has focused on analyzing independent and identically distributed (iid) settings, especially in the horizontal FL (HFL) approach, to achieve better results in terms of convergence, accuracy, communication time, and cost [31]. However, for both federated and centralized data in the context of a DT, these algorithms have yet to be fully considered from a privacy and fairness perspective.

Furthermore, hybrid FL (HFL) has been introduced as another approach that combines HFL and vertical FL (VFL) [33] and shows promise for future work. HFL and VFL each refer to data distribution and require different algorithms. The presence of both data distributions and other federated learning paradigms requires evaluation to determine their potential benefits in terms of convergence time and accuracy. Hetero FL [34] is another approach that allows training a global model based on different local model architectures. This is particularly relevant for logistics DT, where data access is handled at the attribute level, which may be different for each logistics service provider.

Regarding privacy, there are ideas for knowledge transfer [35], with research focusing primarily on public centralized data and transferring knowledge from the locally trained models using one-way knowledge transfer.

Despite these research directions, there are still open issues that need to be addressed. Kairouz et al. [36] examine various challenges associated with federated learning (FL), including the context of cross-silo FL and fairness. The term "cross-silo" refers to the process of training models across multiple organizations, each managing large data silos. These issues include the absence of certain features due to varying data sharing rules and incomplete data entry, which can affect data trust. Other problems are difficulties in data normalization, like different storage formats and inconsistency of data and labels, differences in privacy policies among logistics service providers, and fairness concerns regarding the selection of training participants based on available hardware and further features.

Many of them can be still optimized, validated and further developed in the context of DT environment. In addition, there is a need to explore the combination of the HFL and MFL approaches, which includes the basic distribution of the data to a data trustee in the logistics domain.

III. Prototype

For the prototype, a new access and usage control mechanism will be implemented first to fulfill the requirements of a secure DT with specific requirements in logistics. Once this implementation is complete, the following stage involves requirements engineering, implementation and deployment of the FL framework on a scalable engine, such as Docker [37].

A. Access Control for data trustees in logistics

To meet the requirements of protecting and sharing data at the attribute level, for both internal and external data, the access control model should be designed using state-of-the-art techniques.

An access control mechanism based on ABAC, in more specific Next Generation Access Control (NGAC), should be implemented to secure the internal data and to be able to share it with other logistics service providers and researchers. Through an access control system for FL, researchers only have access to the in- and output of the FL Framework and not directly to the data.

This access control will be implemented as an extension of the Policy Machine [38], in which NGAC is already integrated. This framework can also be used with other access mechanisms such as RBAC. However, fine-grained data sharing at the attribute level, which is essential for logistics, has not yet been integrated.

Figure 1 illustrates the initial design draft of the access data model, which will serve as a repository for the storage of the access rules. In addition to this model, there are four group categories and several further restrictions. These groups are categorized as *standard, FL-internal, FL-external* and *any*. Considering that the platform is designed for logistics, the term "company" is used in the following text to refer to these groups. The *standard* company represents the logistic companies themselves. The *FL-internal* company consists of the platform provider and those who require access to the data for the FL training. They can get it through the FL control access model via a search API, which is only available to specific individuals or for internal data exchange to the FL framework. The *FL-external* company is formed by researchers who are allowed to train AI models by executing jobs through an API, but they are not permitted to have direct access to the data. There is always a mandatory privacy measure for trained models, with minimal privacy requirements for AI models returned to this company. The fourth company is called *any* and consists of all company members who also have access to the *standard* company type.

The user's access is determined by the group to which they belong, with additional consideration given to roles within *standard* and *any* company. If the data come from the user's company or has been shared, there is a second

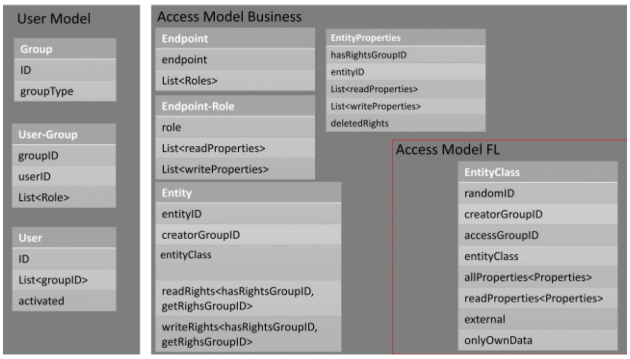

Fig. 1 Access Data Model – initial draft

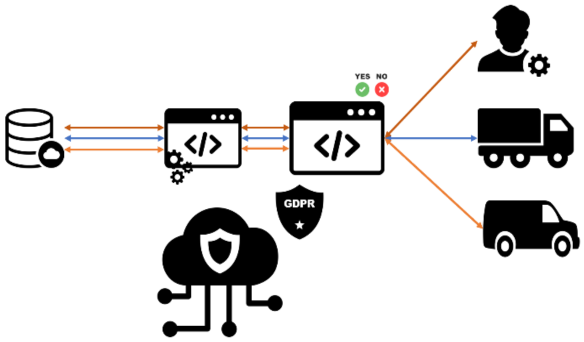

Fig.2 Concept of DT. Icons are from [42].

level of access control. In terms of training, FL algorithms have two levels of access. They can either use only a company's own data, or they can access all shared data from companies. This works according to the permissions granted by the logistics service providers. This two-level access provides better control over data usage and provides an incentive for the platform. The next step involves the transformation in the NGAC access schema and extends it with the attribute level and metadata access control schema.

B. Data trust in logistics

The DT is a web service designed to secure and manage logistics company data. Direct access to the data is restricted by the participating companies and can be shared with other companies or utilized as training input in the FL. An overview of data trust is shown in Fig. 2 and the FL concept is shown in Fig. 3. Fig. 2 shows the companies, the data providers and users and the data trustee, who have access to the data and the processing unit based on the access control rules.

To increase participants' trust in the platform, the DT incorporates state-of-the-art techniques such as Zero Trust Architecture [16] with micro-segmentation, which brings the access control mechanism closer to the data source [17]. The DT also aims to encourage companies to provide their own data for research purposes. The first incentive is that the platform provider acts as a data trustee under a non-profit organization. The second stimulus is the potential benefit of utilizing a federated pricing model or being recommended as a potential collaboration partner if a company shares its data for FL. Another important aspect to consider is fairness in terms of prediction accuracy based on the data provided. The fourth encouragement concerns the data available to researchers, who will only have access to the FL framework and not directly to the data.

For protection against privacy attacks on AI models, any external access to the model should be subject to a mandatory privacy mechanism. These trained models can then be utilized by data scientists and companies for a variety of use cases, such as numerous pricing models or other predictions, such as time optimization or CO2 reduction. There are also ongoing research efforts exploring new data platforms [39], model input parameter selection [40] and pricing strategies in competition between new and existing logistics service providers [41].

Furthermore, depending on the role within the AI ecosystem and whether the data is provided by companies for training or only used by researchers, there should be control over the use of the trained models, as well as the accuracy and privacy of the predictions.

C. FL-DT

Based on the access control model of the metadata for AI training, an FL training framework will be developed. The architectural concept is illustrated in Fig. 3, where companies, the data trustee and researchers have access to the FL training environment, secured by access control. In this example, a company participates in the training with its local data. The infrastructure for the platform and the FL framework is hosted on a scalable framework network.

The first module is data preprocessing. This is where the data are transformed or pre-processed to get the local data in a comparable input format. For this purpose, the data are transferred to a " trusted environment". This environment can be hosted either at the logistics provider, in the platform on the scalable environment, or at a provider trusted by the logistics provider. The logistician decides where this environment exists.

It is like a container that takes care of data preparation and model training. Within the „trusted environment," it is now possible to transform the data, which can be done by the logistician itself or in exchange with the neutral entity of the data trustee, to obtain high quality training data. Additionally, the data can be scaled using a pre-configured and secure scaler. This scaler can be pre-trained solely on the central data from the same company or with privacy additions from all companies. At this point, it should also be possible to tag the imported data with release numbers for reproducibility and explainability. It should also be possible to delete a release if the logistician wishes to do so or regulatory rights require this.

Once this data has been provided, the training phase can commence. This involves addressing various gaps that are specific to DT in logistics and FL. Due to the secure infrastructure approach, all data access and exchange pass

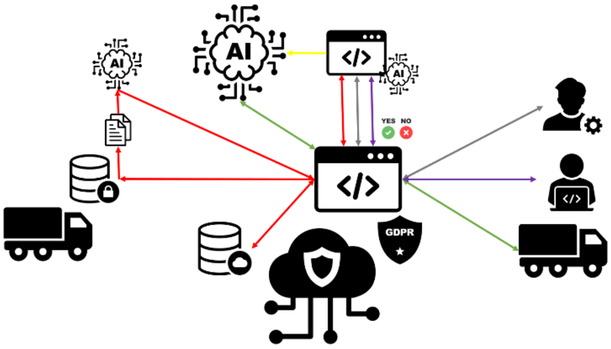

Fig. 3 Concept of FL in data trust. Icons are from [42].

through the access control module, ensuring authentication and authorization. To address this privacy concerns during the training phase, privacy mechanisms such as DP or one-time passwords be employed for the model's data exchange. At this point, blockchain technology would be avoided. This is due to transactional time limitations and legal requirements, such as the right to erasure outlined in the General Data Protection Regulation (GDPR), where only some research idea results are available [43], which are not yet as effective as they need to be.

The model initialization and pretraining is the first step and different approaches exist based on security and fairness considerations. One solution is to pre-train a central model for faster convergence [29], but in the context of DT with sensitive data and no public data. The second approach is to use MFL. The question of which FL approach can help to overcome the challenges posed by different data distributions and sharing combinations in the logistics domain is crucial at this step.

The second step involves training of the model. In this step, the model is transmitted to the "trusted environment" where the pre-processed local data are stored. During the training phase, only the model itself and the training parameters are exchanged with this environment. It is important to implement security measures to ensure that the model works well, while considering the security implications for the participants. One approach to achieve this is the utilization of Differential Privacy (DP).

During training it may be possible to use alternative models, due to different attribute sharing levels between the companies. This could work similarly to transfer learning, whereby the model first trained to the same output error as the central model and then further trained on local data. If this local model works better for their specific use cases, it can be saved and accessed exclusively by the company.

After the training phase, the models are returned to the center where the models are combined by various algorithms or where FL-PATE [35], a knowledge transfer algorithm, would be used. Subsequently, another set of central data is then utilized to train the model.

The next step involves testing whether the model convergence criteria have met. If the criteria are satisfied, the central model is stored in a data repository accessible to logistics service providers for their predictions, such as the delivery price. The accuracy will be different, or a range of output values will be predicted based on the quality and quantity of data supplied by the company for FL training. Researchers can download their trained models based on their chosen configurations or algorithms, with an additional privacy budget, such as DP, to ensure privacy preservation.

IV. Future Work

This paper presents a concept of the FL paradigm that is based on a DT in the logistics domain. Implementing the concept of FL framework in a scalable, secure, fair, and adaptable manner is aimed at in future work. Therefore, the focus would be on the following points that should be considered and researched.

A pre-research topic is to discover pricing models, as discussed in [44], and to find methods to automatically select relevant input attributes based on the secure DT and FL approach.

Furthermore, MFL and pre-trained FL be explored for faster convergence and higher accuracy with secure training. Next, HFL is an important topic for this FL approach. This happens because there are horizontal data across companies. On the other hand, there is also vertical data, such as shared orders, which are provided in diverse ways by the two companies involved in an FL training session. The approach must also be developed so that all shared data can be effectively incorporated into the training of the AI in accordance with the data sharing rules. For the central data in MFL, where the central data is also secret, a secure paradigm for the training process must also be used to ensure secure computation. For this task, it is necessary to be GPDR compliant, e.g., with secure aggregation [24]. Then fairness principles are required, to avoid over-representation of a single company in the AI prediction. Furthermore, exploring alternative incentives for companies to share their data is also essential and how these can be incorporated into the model training.

Considering asynchronous, decentralized, or hierarchical FL approaches in the context of a DT can be beneficial to achieve faster convergence and avoid single points of failure, which is also a part of further research. Finally, a scalable approach should be developed that works in a cluster setup for scalability.

This implementation will be empirically evaluated based on real data from small companies in Saxony, Germany, which, as usual, are interested in receiving an incentive and providing their business data for this purpose.

References

[1] BMDV. “Amtliche Güterkraftverkehrsstatistik.” https://www.kba.de/DE/Statistik/Kraftverkehr/deutscherLastkraftfahrzeuge/vd_Inlandsverkehr/vd_inlandsverkehr_node.html (accessed Apr. 18, 2023).

[2] Eurostat. “Summary of annual road freight transport by type of operation and type of transport (1 000 t, Mio Tkm, Mio Veh-km).” https://

www.eea.europa.eu/ds_resolveuid/ 2952faa0aff24c37aa0cfab6a86730c8 (accessed Apr. 20, 2023).

[3] "A fifth of road freight kilometres by empty vehicles," Eurostat, 12 Oct., 2021. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20211210-1 (accessed: Apr. 20, 2023).

[4] A. Streim and B. Kokott. "In der Logistik werden die Sicherheitsmaßnahmen verschärft." https://www.bitkom.org/Presse/Presseinformation/Digitalisierung-Logistik (accessed Apr. 14, 2023).

[5] C. Kille, T. Schmidt, W. Stölzle, L. Häberle, and S. Rank. "Begegnung von Kapazitätsengpässen im Straßengüterverkehr – Fokus Personal." http://logistik-digitalisierung.de/ (accessed Apr. 18, 2023).

[6] W. Hall and J. Pesenti, "Growing the artificial intelligence industry in the UK," 2017. [Online]. Available: https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk

[7] [7] Insititut für Wirtschaftsinformatik – Universität Leipzig, TRANSIT Data trust for logistics: Data Trusts for Enhancing Logistics Collaboration. Accessed: May 22, 2023. [Online]. Available: https://transit-project.de/

[8] J. Witkowski, "Electronic Freight exchange and logsitics platforms in buildung of supply chains," 2018. [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi734WPh8L-AhXQp6QKHSo9BdwQFnoECAgQAQ&url=https%3A%2F%2Fwww.confer.cz%2Fclc%2F2018%2Fdownload%2F2448-european-electronic-freight-exchange-as-a-future-central-coordinator-in-supply-chains.pdf&usg=AOvVaw38zq7w-pkr1klk3gvC12Vu

[9] K. Houser and J. W. Bagby, The Data Trust Solution to Data Sharing Problems, 2022.

[10] B. Liu, N. Lv, Y. Guo, and Y. Li, "Recent Advances on Federated Learning: A Systematic Survey," Jan. 2023, http://dx.doi.org/10.48550/arXiv.2301.01299. [Online]. Available: http://arxiv.org/pdf/2301.01299v1

[11] S. Stalla-Bourdillon, G. Thuermer, J. Walker, L. Carmichael, and E. Simperl, "Data protection by design: Building the foundations of trustworthy data sharing," Data & Policy, vol. 2, 2020, http://dx.doi.org/10.1017/dap.2020.1.

[12] X. Zhang, "A commentary of Data trusts in MIT Technology Review 2021," Fundamental Research, vol. 1, no. 6, pp. 834–835, 2021, http://dx.doi.org/10.1016/j.fmre.2021.11.016.

[13] R. K. Lomotey, S. Kumi, and R. Deters, "Data Trusts as a Service: Providing a platform for multi-party data sharing," International Journal of Information Management Data Insights, vol. 2, no. 1, p. 100075, 2022, http://dx.doi.org/10.1016/j.jjimei.2022.100075.

[14] A. Blankertz. "Designing Data Trusts: Why We Need to Test Consumer Data Trusts Now." https://www.stiftung-nv.de/sites/default/files/designing_data_trusts_e.pdf (accessed Apr. 20, 2023).

[15] C. L. Geminn, P. C. Johannes, J. K. M. Müller, and M. Nebel, Data Governance in Germany – An Introduction. Universität Kassel, 2023. [Online]. Available: https://kobra.uni-kassel.de/handle/123456789/14590

[16] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture," National Institute of Standards and Technology, 2020. http://dx.doi.org/10.6028/NIST.SP.800-207. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-207/final?ref=hackernoon.com

[17] Ruoming Pang et al., "Zanzibar: {Google's} Consistent, Global Authorization System," in Proceedings of the 2019 USENIX Annual Technical Conference: July 10-12, 2019, Renton, WA, USA, 2019, pp. 33–46. [Online]. Available: https://www.usenix.org/conference/atc19/presentation/pang

[18] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," http://dx.doi.org/10.48550/arXiv.1602.05629. [Online]. Available: http://arxiv.org/pdf/1602.05629v4

[19] N. F. Syed, S. W. Shah, A. Shaghaghi, A. Anwar, Z. Baig, and R. Doss, "Zero Trust Architecture (ZTA): A Comprehensive Survey," IEEE Access, vol. 10, pp. 57143–57179, 2022, http://dx.doi.org/10.1109/access.2022.3174679.

[20] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," Computer, vol. 29, no. 2, pp. 38–47, 1996, http://dx.doi.org/10.1109/2.485845.

[21] D. F. Ferraiolo, R. Chandramouli, V. C. Hu, and D. R. R. Kuhn, "A Comparison of Attribute Based Access Control (ABAC) Standards for Data Service Applications," 2016, http://dx.doi.org/10.6028/NIST.SP.800-178.

[22] Ruoming Pang et al., "Zanzibar: Google's Consistent, Global Authorization System," 2019. [Online]. Available: https://www.semanticscholar.org/paper/Zanzibar%3A-Google's-Consistent%2C-Global-Authorization-Pang-C%C3%A1ceres/1362dec32d9d0b9d8b369f7ebcfef19bbc975066

[23] R. Sandhu and J. Park, "Usage Control: A Vision for Next Generation Access Control," in vol. 2776, 2003, pp. 17–31, http://dx.doi.org/10.1007/978-3-540-45215-7_2.

[24] N. Truong, K. Sun, S. Wang, F. Guitton, and Y. Guo, "Privacy Preservation in Federated Learning: An insightful survey from the GDPR Perspective," Nov. 2020, http://dx.doi.org/10.48550/arXiv.2011.05411. [Online]. Available: https://arxiv.org/pdf/2011.05411

[25] J. Zhang, H. Zhu, F. Wang, J. Zhao, Q. Xu, and H. Li, "Security and Privacy Threats to Federated Learning: Issues, Methods, and Challenges," Security and Communication Networks, vol. 2022, pp. 1–24, 2022, http://dx.doi.org/10.1155/2022/2886795.

[26] M. Alazab, S. P. RM, P. M, P. K. R. Maddikunta, T. R. Gadekallu, and Q.-V. Pham, "Federated Learning for Cybersecurity: Concepts, Challenges, and Future Directions," IEEE Trans. Ind. Inf., vol. 18, no. 5, pp. 3501–3509, 2022, http://dx.doi.org/10.1109/TII.2021.3119038.

[27] J. Zhu, J. Cao, D. Saxena, S. Jiang, and H. Ferradi, "Blockchain-empowered Federated Learning: Challenges, Solutions, and Future Directions," ACM Comput. Surv., vol. 55, no. 11, pp. 1–31, 2023, http://dx.doi.org/10.1145/3570953.

[28] G. Yu et al., "IronForge: An Open, Secure, Fair, Decentralized Federated Learning," Jan. 2023, http://dx.doi.org/10.48550/arXiv.2301.04006. [Online]. Available: https://arxiv.org/pdf/2301.04006

[29] J. Nguyen, J. Wang, K. Malik, M. Sanjabi, and M. Rabbat, "Where to Begin? On the Impact of Pre-Training and Initialization in Federated Learning," Oct. 2022, http://dx.doi.org/10.48550/arXiv.2210.08090. [Online]. Available: https://arxiv.org/pdf/2210.08090

[30] S. Reddi et al., "Adaptive Federated Optimization," Feb. 2020, http://dx.doi.org/10.48550/arXiv.2003.00295. [Online]. Available: https://arxiv.org/pdf/2003.00295

[31] S. Augenstein et al., "Mixed Federated Learning: Joint Decentralized and Centralized Learning," May. 2022, http://dx.doi.org/10.48550/arXiv.2205.13655. [Online]. Available: https://arxiv.org/pdf/2205.13655

[32] K. Yang, S. Chen, and C. Shen, "On the Convergence of Hybrid Server-Clients Collaborative Training," IEEE J. Select. Areas Commun., vol. 41, no. 3, pp. 802–819, 2023, http://dx.doi.org/10.1109/JSAC.2022.3229443.

[33] X. Zhang, W. Yin, M. Hong, and T. Chen, "Hybrid Federated Learning: Algorithms and Implementation," Dec. 2020, http://dx.doi.org/10.48550/arXiv.2012.12420. [Online]. Available: https://arxiv.org/pdf/2012.12420

[34] E. Diao, J. Ding, and V. Tarokh, "HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients," Oct. 2020, http://dx.doi.org/10.48550/arXiv.2010.01264. [Online]. Available: https://arxiv.org/pdf/2010.01264

[35] Y. Pan, J. Ni, and Z. Su, "FL-PATE: Differentially Private Federated Learning with Knowledge Transfer," in 2021 IEEE Global Communications Conference (GLOBECOM), 2021, http://dx.doi.org/10.1109/globecom46510.2021.9685079.

[36] P. Kairouz et al., "Advances and Open Problems in Federated Learning," Dec. 2019, http://dx.doi.org/10.48550/arXiv.1912.04977. [Online]. Available: http://arxiv.org/pdf/1912.04977v3

[37] Dirk Merkel, Docker: lightweight linux containers for consistent development and deployment. Houston, TX: Belltown Media, 2014. [Online]. Available: https://dl.acm.org/doi/10.5555/2600239.2600241

[38] D. Ferraiolo, V. Atluri, and S. Gavrila, "The Policy Machine: A novel architecture and framework for access control policy specification and enforcement," Journal of Systems Architecture, vol. 57, no. 4, pp. 412–424, 2011, http://dx.doi.org/10.1016/j.sysarc.2010.04.005.

[39] Y.-A. Du, "Research on the Route Pricing Optimization Model of the Car-Free Carrier Platform Based on the BP Neural Network Algorithm," Complexity, vol. 2021, pp. 1–10, 2021, http://dx.doi.org/10.1155/2021/8204214.

[40] M. Poliak, A. Poliakova, L. Svabova, N. A. Zhuravleva, and E. Nica, "Competitiveness of Price in International Road Freight Transport,"

JOC, vol. 13, no. 2, pp. 83–98, 2021, http://dx.doi.org/10.7441/joc.2021.02.05.

[41] F. Du, S. Ang, F. Yang, and C. Yang, "Price and distribution range of logistics service providers considering market competition," APJML, vol. 30, no. 4, pp. 762–778, 2018, http://dx.doi.org/10.1108/APJML-09-2017-0208.

[42] UXWing.com, Exclusive collection of free icons download for commercial projects without attribution. [Online]. Available: https://uxwing.com

[43] E. Politou, F. Casino, E. Alepis, and C. Patsakis, "Blockchain Mutability: Challenges and Proposed Solutions," Jul. 2019, http://dx.doi.org/10.48550/arXiv.1907.07099. [Online]. Available: https://arxiv.org/pdf/1907.07099

[44] H.-S. Jang, T.-W. Chang, and S.-H. Kim, "Prediction of Shipping Cost on Freight Brokerage Platform Using Machine Learning," Sustainability, vol. 15, no. 2, p. 1122, 2023. http://dx.doi.org/10.3390/su15021122. [Online]. Available: https://www.mdpi.com/2071-1050/15/2/1122