Leveraging the Potentials of Federated AI Ecosystems

Marco Röder
University of Würzburg, Chair of Information Systems Engineering, Würzburg, Germany

Peter Kowalczyk
University of Würzburg, Chair of Information Systems Engineering, Würzburg, Germany

Frédéric Thiesse
University of Würzburg, Chair of Information Systems Engineering, Würzburg, Germany

Leveraging the Potentials of Federated AI Ecosystems

Marco Röder[1], Peter Kowalczyk[1], and Frédéric Thiesse[1]

[1] University of Würzburg, Chair of Information Systems Engineering, Würzburg, Germany
{marco.roeder,peter.kowalczyk,frederic.thiesse}@uni-wuerzburg.de

Abstract. Deep learning increasingly receives attention due to its ability to efficiently solve various complex prediction tasks in organizations. It is therefore not surprising that more and more business processes are supported by deep learning. With the proliferation of edge intelligence, this trend will continue and, in parallel, new forms of internal and external cooperation are provided through federated learning. Hence, companies must deal with the potentials and pitfalls of these technologies and decide whether to deploy them or not and how. However, there currently is no domain-spanning decision framework to guide the efficient adoption of these technologies. To this end, the present paper sheds light on this research gap and proposes a research agenda to foster the potentials of value co-creation within federated AI ecosystems.

Keywords: Edge Intelligence, Federated Learning, AI Ecosystem

1 Introduction

The umbrella term “deep learning” (DL) denotes algorithms from the broader field of artificial intelligence (AI) that seek to train complex artificial neural networks, which typically consist of numerous layers between the model input and output [1, 2]. Such deep neural networks (DNN) are particularly suited to process vast amounts of data effectively to solve prediction tasks [3]. Thus, DL holds the potential to drive a wide range of processes in important corporate areas, such as fraud detection, decision support, automation, and more [2, 4, 5]. However, the application of DL is also accompanied by challenges like learning from sparse data, model bias, poor model performance, or maintaining data privacy [6–8].

In light of the advances in cloud-based systems, DL components are increasingly used for business tasks as mentioned before [9, 10]. Moreover, a study by Deloitte from 2019 indicates that Internet of Things (IoT) projects using AI technologies will increase by 70 percent until 2022 [11]. With the proliferation of edge intelligence (EI) technologies, which push DL towards the edge of the network (e.g., IoT-devices, and edge servers), this distribution trend of DNN is continued [12]. Additionally, EI enables new collaboration potentials at various organizational stages by utilizing federated learning (FL) [13]. The objective of FL is to train a shared global DNN with the insights gained from decentral DNNs instantiated by locally dispersed clients [14, 15].

For example, by deploying EI, an electronic article surveillance (EAS) system in retail (e.g., as proposed by Hauser et al. [16]), could be extended to facilitate DNNs on

16th International Conference on Wirtschaftsinformatik,
March 2021, Essen, Germany

local EI devices such as, for example, RFID gates in stores. If FL is applied additionally, the local DNN could be trained collaboratively with insights gathered from other RFID gates located in the same store, with those from EAS systems in a larger retail store network, or even jointly with company-external sources.

Drawing on recent literature on ecosystems further substantiates the idea of such an interwoven application of EI and FL to build more sophisticated DL models. The term ecosystem originates from biology and is generally referred to as the fusion of multiple units that interact with each other and the environment [17, 18]. As far as data ecosystems are concerned, the ecosystem units share data either intra- or inter-organizational [19]. With regard to a federated AI ecosystem, shared insights from the EI instances (i.e., the entities of the ecosystem) can be either related to a specific task or even to integrated processes. The more entities involved in such a federated AI ecosystem, the greater the chance and possible magnitude of benefit for each of them [20, 21]. Thus, we leverage these possible effects by taking the ecosystem perspective [22] and loosely following the service-dominant (SD) logic put forward by Vargo and Lusch [23, 24], which emphasizes services (i.e., intangibles) rather than goods (i.e., tangibles) as the resources of exchange to co-create value [23, 25–27].

Combining EI and FL holds the potential to enhance the system's performance, generalizability and robustness, and thus assist to overcome current challenges associated with AI in practice (i.e., model bias, sparse data, data privacy, poor model performance) [28–32]. However, while current research endeavors are already directed towards the development of specific systems deploying FL [33, 34]—to the best of our knowledge—there is no guidance on how to identify and enhance suitable processes to leverage the potentials of EI and FL for value co-creation in ecosystems. To this end, we propose our research question as follows: *How can FL be used to empower AI ecosystems for value co-creation?*

In the following sections, we first elaborate on the technological background of EI and FL. We then present a corresponding research agenda to serve as a blueprint to assist and motivate researchers as well as practitioners to engage with this promising topic. Subsequently, we conclude the present paper by applying the design-oriented research methodology (DSRM) as proposed by Peffers et al. [35] to the research agenda and briefly outline the expected contributions.

2 Theoretical Background

2.1 Edge Intelligence

EI follows the edge computing (EC) paradigm [12]. EC can be described as a distributed and decentralized computing concept [36] which enables data processing to happen directly or in proximity to the data source [37]. More specifically, EC includes all nodes along the path from the end devices (e.g., sensors), over edge servers (e.g., micro-data centers) to the cloud data center [37]. For the sake of simplicity, we generally refer to these points as "edge nodes" (EN). Now, EI (cf. Figure 1) can be regarded as the migration of traditionally cloud-based DNN to these ENs [34, 38]. Therefore, EI can

overcome the specific issues associated with cloud computing (e.g., latency, data privacy, or communication inefficiency) [12, 37, 39–41]. Furthermore, shifting data processing to the edge of the network makes transferring all raw data to a central cloud unit obsolete [42]. Instead, data processing can take place in closer proximity to its origin, and thus preprocessed data are transferred [12, 42]. Each EN in this EI hierarchy is capable of consuming and producing data (e.g., by inferencing) [12]. Following the definitions of Zhou et al. and Xu et al., we refer to EI as the usage of AI algorithms locally on any of the ENs to enhance model training and inferencing, while simultaneously protecting the privacy and security of data [12, 42]. According to the idea of EC, each EN in this hierarchy is capable of collaborating with other nodes vertically or horizontally [12].

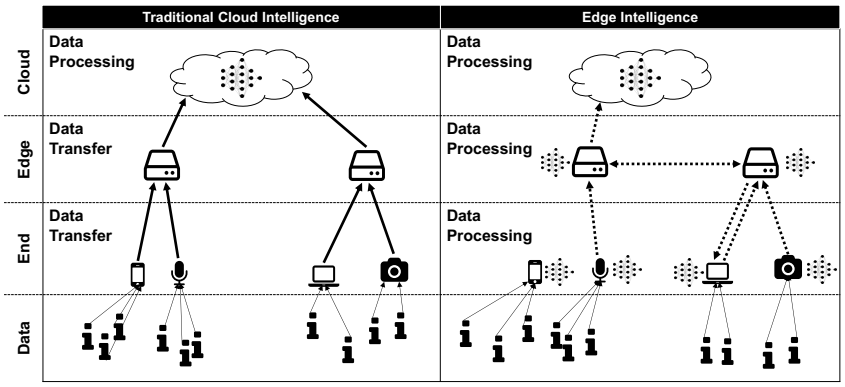

Figure 1. Comparison of traditional cloud intelligence and edge intelligence [12, 42]

2.2 Federated Learning

In order to facilitate vertical or horizontal collaborative training of distributed DNNs, FL poses a promising solution [12]. The objective of FL is to train a shared global DL model provided by a high-level instance (model owner) by successively feeding insights gained from decentral DNNs which are instantiated by locally dispersed clients (data owners) [12–14, 34]. Therefore, the local DNN iteratively updates the global model [14]. Here it should be emphasized that private data are treated confidentially in the sense that they are not forwarded but rather remain with the data owner [12, 14]. Instead, only the parameter values of the local DNNs are used to update the global DNN, ideally making plausible data protection concerns obsolete [12, 34]. The training procedure of FL (cf. Table 1) can be divided into three steps: (1) task initialization, (2) local model training and updating, (3) global model aggregation and updating [34].

Although this decentralized learning approach is rich in potential (i.e., privacy protection, reduction of model bias), FL may also come along with downsides—namely algorithmic or practical challenges [28]. While the former may emerge by the difficulty to design an appropriate model averaging policy that is fast and robust despite limited availability of model updates or malicious contributors, the latter results from practical issues such as the restorability of private data by another client [28, 33, 43].

Table 1. Steps of federated learning [34]

Step		Description
	(1)	The model owner decides upon the training task and necessary data, initializes model hyperparameters, and shares the initialized model (G_i) with the data owners.
Repeat	(2)	Each data owner applies G_i (or G_{i+j} respectively) as a local model and optimizes this model with private data. Finally, the data owner sends the updated local model parameters back to the model owner.
	(3)	The model owner receives the updated parameters from the data owners and aggregates these updates effectively to a new global model (G_{i+j}). G_{i+j} is then sent back to the data owners.

3 Research Agenda

As illustrated in section two, EI can lead to a reduction in latency, improves communication efficiency, and increases data security [12, 37, 39–41]. Additionally, FL may potentially help to overcome some of the hurdles in the context of AI deployment (i.e., model bias, sparse data, model performance) [28–32]. By combining both technologies we merge advantages and opt for a system which delivers a secure and efficient communication of the necessary information to build more sophisticated DL models in terms of performance, generalizability, and robustness. Now, by taking the SD logic perspective, we argue that building service ecosystems, which incorporate these technologies and additionally connect multiple entities, resembles a promising research field to be investigated further. Therefore, we encourage researchers and practitioners to engage with federated AI ecosystems by working on the following questions:

- Which processes can be enhanced by EI technologies and provide the potential for value co-creation based on FL through the exchange of insights?
- How to design and operate an effective SD platform with a reasonable modular FL architecture at company level?
- How to configure, monitor, and manage a federated AI ecosystem at an inter-company level to leverage the full potentials of value co-creation?
- How to maintain data security and prevent the recovery of original data in federated AI ecosystems?

4 Future Work and Expected Contribution

In the light of the identified research gap and our proposed agenda, we encourage researchers and practitioners to engage with this topic. Against this backdrop, we propose three possible follow-up studies that are directly associated with the aforementioned research agenda. Here, we especially focus on the first study and outline its backbone in depth (cf. Table 2). To this end, we follow the DSRM approach put forward by Peffers et al. [35]. Briefly summarized, design science research—besides behavioral research—as one of the two pillars of IS research offers a methodological toolset to create useful artifacts which are often directed towards business contexts [44–47].

Table 2. Overview on study 1, in line with the DSRM [35]

Identify problem and motivate	Configuring FL models to facilitate value co-creation in business networks and therefore outperform local instantiations due to generalizability and robustness remains an unexplored potential for a wide range of business applications. These circumstances determine the entry point of this first study.
Define objectives and solution	We attribute this lack of practical value co-creation solutions to the absence of a corresponding decision framework that determines a suitable configuration of EI and FL for the specific task under consideration.
Design and development	Hence, an artifact is designed to (i) identify processes to be enhanced with EI and FL, and (ii) to guide the effective implementation of such technologies to leverage the potentials of value co-creation. The decision framework is therefore not restricted to specific application domains, edge devices nor DNN configurations.
Demonstration	Given a real-world application scenario with its corresponding environment of stakeholders, we aim for a first demonstration of the novel artifacts' utility.
Evaluation	The evaluation is carried out in a formative and naturalistic manner [48]. More precisely, we aim for a stepwise assessment of the artifact's effectiveness in a real-world application scenario.
Communication	The core of this first study is the development of a decision framework for the identification and enhancement of processes with EI and FL. The research findings are communicated via journals and conference proceedings.

Drawing on the results of the first study (i.e., the decision framework), a consecutive study aims to assist companies with regard to the adoption of suitable FL models. To this end, we develop a service platform with the capability to accumulate insights from locally dispersed entities in a FL model to empower multiple corporate-specific processes with DL. Again, we plan to opt for a design-oriented research approach to develop the platform solution while considering its stakeholder's requirements.

A third and last proposed study extends the idea of a service platform by taking the inter-company perspective. Thus, the participating clients form a service ecosystem to share and therefore improve the robustness and generalizability of the FL model across multiple companies. Additionally, new ecosystem attendees benefit from the guided adoption of sophisticated DL models. For the purpose of control and enhancement, suitable metrics and components to real-time monitor and benchmark such a service ecosystem (e.g., in terms of latency or performance) are incorporated.

This article set out to propose the idea of federated AI ecosystems by merging both technologies EI and FL and by taking the ecosystems perspective. Furthermore, we elaborated a research agenda to boost the discussion in the IS community. Ultimately, we sketched out three possible follow-up studies at the nexus between EI, FL, and the SD logic perspective by applying the DSRM. However, as the research agenda shows, more research is yet to be conducted in this area.

References

1. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural Networks. 61, 85–117 (2015).
2. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature. 521, 436–444 (2015).
3. Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.: Traffic Flow Prediction With Big Data: A Deep Learning Approach. IEEE Trans. Intell. Transp. Syst. 16, 865–873 (2015).
4. Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E.: Deep learning applications and challenges in big data analytics. J. Big Data. 2, 1 (2015).
5. Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., Beling, P.: Deep learning detecting fraud in credit card transactions. In: 2018 Systems and Information Engineering Design Symposium (SIEDS). pp. 129–134. IEEE, Charlottesville, VA (2018).
6. Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E.: Deep learning applications and challenges in big data analytics. J. Big Data. (2015).
7. van der Aalst, W.M.P., Bichler, M., Heinzl, A.: Responsible Data Science. Bus. Inf. Syst. Eng. 59, 311–313 (2017).
8. Martin, K.E.: Ethical issues in the big data industry. MIS Q. Exec. 14, 67–85 (2015).
9. Li, H., Ota, K., Dong, M.: Learning IoT in Edge: Deep Learning for the Internet of Things with Edge Computing. IEEE Netw. 32, 96–101 (2018).
10. Huang, Y., Ma, X., Fan, X., Liu, J., Gong, W.: When deep learning meets edge computing. In: Proceedings of the 25th International Conference on Network Protocols (ICNP). pp. 1–2. IEEE (2017).
11. Deloitte: Bringing AI to the device: Edge AI chips come into their own. (2019).
12. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., Zhang, J.: Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. Proc. IEEE. 107, 1738–1762 (2019).
13. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated Machine Learning : Concept and Applications. ACM Trans. Intell. Syst. Technol. 10, 1–19 (2019).
14. Brendan McMahan, H., Moore, E., Ramage, D., Hampson, S., Agüera y Arcas, B.: Communication-efficient learning of deep networks from decentralized data. Proc. 20th Int. Conf. Artif. Intell. Stat. AISTATS 2017. 54, (2017).
15. Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D.: Federated Learning: Strategies for Improving Communication Efficiency. 1–10 (2016).
16. Hauser, M., Zügner, D., Flath, C., Thiesse, F.: Pushing the limits of RFID: Empowering RFID-based electronic article surveillance with data analytics techniques. In: Proceedings of the International Conference on Information Systems (ICIS) (2015).
17. Moore, J.F.: The Death of Competition: Leadership and Strategy in the Age of

Business Ecosystems. (1996).
18. Chapin, F.S., Matson, P.A., Mooney, H.A.: Principles of Terrestrial Ecosystem Ecology. (2002).
19. Lis, D., Otto, B.: Data Governance in Data Ecosystems - Insights from Organizations. In: Proceedings of Americas' Conference on Information Systems (AMCIS) (2020).
20. Jarke, M., Otto, B., Ram, S.: Data Sovereignty and Data Space Ecosystems. Bus. Inf. Syst. Eng. 61, 549–550 (2019).
21. Gelhaar, J., Otto, B.: Challenges in the emergence of data ecosystems. In: Proceedings of the 24th Pacific Asia Conference on Information Systems (PACIS) (2020).
22. Shipilov, A., Gawer, A.: Integrating research on interorganizational networks and ecosystems. Acad. Manag. Ann. (2020).
23. Vargo, S.L., Lusch, R.F.: Evolving to a New Dominant Logic for Marketing. J. Mark. (2004).
24. Vargo, S.L., Lusch, R.F.: Service-dominant logic: Continuing the evolution. J. Acad. Mark. Sci. 36, 1–10 (2008).
25. Akaka, M.A., Vargo, S.L., Lusch, R.F.: The complexity of context: A service ecosystems approach for international marketing. J. Int. Mark. 21, 1–20 (2013).
26. Lusch, R.F., Nambisan, S.: Service innovation: A service-dominant logic perspective. MIS Q. Manag. Inf. Syst. (2015).
27. Vargo, S.L., Lusch, R.F.: Service-dominant logic 2025. Int. J. Res. Mark. (2017).
28. Kairouz, P., Brendan McMahan, H., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., Rouayheb, S. El, Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konečný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S.U., Sun, Z., Suresh, A.T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F.X., Yu, H., Zhao, S.: Advances and open problems in federated learning, (2019).
29. McMahan, H.B., Ramage, D., Talwar, K., Zhang, L.: Learning differentially private recurrent language models. In: Proceedings of the 6th International Conference on Learning Representations (ICLR) (2018).
30. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., Shmatikov, V.: How to backdoor federated learning. arXiv. (2018).
31. Greenland, S., Mansournia, M.A., Altman, D.G.: Sparse data bias: A problem hiding in plain sight. BMJ. (2016).
32. Wang, X., Han, Y., Wang, C., Zhao, Q., Chen, X., Chen, M.: In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning. IEEE Netw. (2019).
33. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., Brendan McMahan, H., Van

Overveldt, T., Petrou, D., Ramage, D., Roselander, J.: Towards federated learning at scale: System design, (2019).
34. Lim, W.Y.B., Luong, N.C., Hoang, D.T., Jiao, Y., Liang, Y.-C., Yang, Q., Niyato, D., Miao, C.: Federated Learning in Mobile Edge Networks: A Comprehensive Survey. IEEE Commun. Surv. Tutorials. 22, 2031–2063 (2019).
35. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. 24, 45–77 (2007).
36. Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., Zomaya, A.Y.: Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. IEEE Internet Things J. 4662, 1–1 (2020).
37. Shi, W., Cao, J., Zhang, Q., Li, Y., Xu, L.: Edge Computing: Vision and Challenges. IEEE Internet Things J. 3, 637–646 (2016).
38. Zhang, X., Wang, Y., Lu, S., Liu, L., Xu, L., Shi, W.: OpenEI: An open framework for edge intelligence. In: Proceedings of the International Conference on Distributed Computing Systems. pp. 1840–1851. IEEE (2019).
39. Mach, P., Becvar, Z.: Mobile Edge Computing: A Survey on Architecture and Computation Offloading. IEEE Commun. Surv. Tutorials. 19, 1628–1656 (2017).
40. Porambage, P., Okwuibe, J., Liyanage, M., Ylianttila, M., Taleb, T.: Survey on Multi-Access Edge Computing for Internet of Things Realization. IEEE Commun. Surv. Tutorials. 20, 2961–2991 (2018).
41. Greengard, S.: AI on Edge. Commun. ACM. 63, 18–20 (2020).
42. Xu, D., Li, T., Li, Y., Su, X., Tarkoma, S., Jiang, T., Crowcroft, J., Hui, P.: Edge Intelligence: Architectures, Challenges, and Applications. (2020).
43. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Process. Mag. (2020).
44. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decis. Support Syst. 15, 251–266 (1995).
45. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. Manag. Inf. Syst. 28, 75–105 (2004).
46. Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. MIS Q. Manag. Inf. Syst. 37, 337–356 (2013).
47. Gregor, S., Jones, D.: The anatomy of a design theory. J. Assoc. Inf. Syst. 8, 312–335 (2007).
48. Venable, J., Pries-Heje, J., Baskerville, R.: FEDS: A Framework for Evaluation in Design Science Research. Eur. J. Inf. Syst. 25, 77–89 (2016).